

Curs 9

Cristian Niculescu

1 Repartiții beta

1.1 Scopurile învățării

1. Să se familiarizeze cu familia cu 2 parametri a repartițiilor beta și cu normalizarea ei.
2. Să poată să actualizeze o a priori beta la o a posteriori beta în cazul unei verosimilități binomiale.

1.2 Repartiția beta

Repartiția beta $\text{beta}(a, b)$ este o repartiție cu 2 parametri cu domeniul $[0, 1]$ și pdf

$$f(\theta) = \frac{(a+b-1)!}{(a-1)!(b-1)!} \theta^{a-1} (1-\theta)^{b-1}.$$

Următoarea aplicație ne permite să explorăm forma repartiției Beta când parametrii variază:

<http://mathlets.org/mathlets/beta-distribution/>.

Repartiția beta poate fi definită pentru orice numere reale $a > 0$ și $b > 0$. Am definit-o doar pentru $a, b \in \mathbb{N}^*$, dar puteți vedea întreaga istorie aici:

http://en.wikipedia.org/wiki/Beta_distribution.

În contextul actualizării bayesiene, a și b sunt numiți adesea **hiperparametri** pentru a-i deosebi de parametrul necunoscut θ care reprezintă ipoteza noastră. Într-un anumit sens, a și b sunt la "un nivel mai sus" decât θ , deoarece ei parametrizează pdf.

1.2.1 O observație simplă, dar importantă

Dacă o pdf $f(\theta)$ are forma $c\theta^{a-1}(1-\theta)^{b-1}$, atunci $f(\theta)$ este pdf a unei repartiții $\text{beta}(a, b)$ și constanta de normalizare trebuie să fie

$$c = \frac{(a+b-1)!}{(a-1)!(b-1)!}.$$

Aceasta rezultă deoarece constanta c trebuie să normalizeze pdf pentru a avea probabilitatea totală 1. Există doar o asemenea constantă și ea este dată în formula pentru repartiția beta.

O observație similară este valabilă pentru repartițiile normală, exponențială, etc.

1.2.2 A priori și a posteriori beta pentru variabile aleatoare binomiale

Exemplul 1. Presupunem că avem o monedă cu probabilitate necunoscută θ a aversului. Aruncăm moneda de 12 ori și obținem 8 aversuri și 4 reversuri. Plecând cu o a priori plată, arătați că pdf a posteriori este o repartiție beta(9, 5).

Răspuns. Notăm cu x_1 datele din cele 12 aruncări. În tabelul următor numim c_2 factorul constant din coloana a posteriori. Observația noastră simplă ne va spune că acesta trebuie să fie factorul constant din pdf beta.

Datele sunt 8 aversuri și 4 reversuri. Deoarece acestea vin dintr-o repartiție binomială(12, θ), verosimilitatea este $p(x_1|\theta) = C_{12}^8 \theta^8 (1 - \theta)^4$. Astfel, tabelul de actualizare Bayesiană este

| hypothesis | prior | likelihood | Bayes numerator | posterior |
|------------|-------------------|---|--|---------------------------------------|
| θ | $1 \cdot d\theta$ | $\binom{12}{8} \theta^8 (1 - \theta)^4$ | $\binom{12}{8} \theta^8 (1 - \theta)^4 d\theta$ | $c_2 \theta^8 (1 - \theta)^4 d\theta$ |
| total | 1 | | $T = \binom{12}{8} \int_0^1 \theta^8 (1 - \theta)^4 d\theta$ | 1 |

Observația noastră simplă de mai sus are loc cu $a = 9$ și $b = 5$. De aceea pdf a posteriori

$$f(\theta|x_1) = c_2 \theta^8 (1 - \theta)^4$$

are o repartiție beta(9, 5) și constanta de normalizare c_2 trebuie să fie

$$c_2 = \frac{13!}{8! \cdot 4!}.$$

Notă. Am inclus explicit coeficientul binomial C_{12}^8 în verosimilitate. Puteam să-l notăm cu c_1 și să nu-i dăm valoarea explicită.

Exemplul 2. Acum presupunem că aruncăm aceeași monedă din nou, obținând n aversuri și m reversuri. Folosind pdf a posteriori din exemplul precedent ca noua noastră pdf a priori, arătați că noua pdf a posteriori este cea a unei repartiții beta(9 + n , 5 + m).

Răspuns. Totul este în tabel. Vom numi x_2 datele acestor $n + m$ aruncări adiționale. De această dată nu vom mai face explicit coeficientul binomial. În loc de aceasta îl vom nota cu c_3 . Ori de câte ori avem nevoie de o nouă

etichetă vom folosi c cu un nou indice. În loc de "Bayes posterior" se va citi "numărătorul Bayes", iar în loc de "numerator" se va citi "a posteriori".

| hyp. | prior | likelihood | Bayes posterior | numerator |
|----------|---------------------------------------|--|---|---|
| θ | $c_2 \theta^8 (1 - \theta)^4 d\theta$ | $c_3 \theta^n (1 - \theta)^m$ | $c_2 c_3 \theta^{n+8} (1 - \theta)^{m+4} d\theta$ | $c_4 \theta^{n+8} (1 - \theta)^{m+4} d\theta$ |
| total | 1 | $T = \int_0^1 c_2 c_3 \theta^{n+8} (1 - \theta)^{m+4} d\theta$ | | 1 |

Din nou observația noastră simplă are loc și, de aceea, pdf a posteriori

$$f(\theta|x_1, x_2) = c_4 \theta^{n+8} (1 - \theta)^{m+4}$$

este cea a unei repartiții beta($n + 9, m + 5$).

Observație. **Beta plată.** Repartiția beta(1,1) coincide cu repartiția uniformă pe $[0, 1]$, pe care am numit-o de asemenea a priori plată pentru θ . Aceasta rezultă înlocuind $a = 1$ și $b = 1$ în definiția repartiției beta, dând $f(\theta) = 1$.

Rezumat. Dacă probabilitatea aversului este θ , numărul de aversuri în $n+m$ aruncări are o repartiție binomială($n+m, \theta$). Am văzut că dacă a priori pentru θ este o repartiție beta, atunci la fel este și a posteriori; doar parametrii a și b ai repartiției beta se schimbă! Rezumăm precis cum se schimbă într-un tabel. Presupunem că datele sunt n aversuri în $n + m$ aruncări.

| hypothesis | data | prior | likelihood | posterior |
|------------|---------|---|-------------------------------|---|
| θ | $x = n$ | beta(a, b) | binomial($n + m, \theta$) | beta($a + n, b + m$) |
| θ | $x = n$ | $c_1 \theta^{a-1} (1 - \theta)^{b-1} d\theta$ | $c_2 \theta^n (1 - \theta)^m$ | $c_3 \theta^{a+n-1} (1 - \theta)^{b+m-1} d\theta$ |

1.2.3 A priori conjugate

Repartiția beta este numită o **a priori conjugată** pentru repartiția binomială. Aceasta înseamnă că, dacă funcția de verosimilitate este binomială, atunci o a priori beta dă o a posteriori beta. De fapt, repartiția beta este o a priori conjugată și pentru repartițiile Bernoulli și geometrică.

Un alt exemplu important: repartiția normală își este propria a priori conjugată. În particular, dacă funcția de verosimilitate este normală cu dispersie cunoscută, atunci o a priori normală dă o a posteriori normală.

A priori conjugate sunt utile deoarece reduc actualizarea Bayesiană la modificarea parametrilor repartiției a priori (așa numiții hiperparametri) în locul calculului integralelor. Am văzut asta pentru repartiția beta în ultimul tabel. Pentru mult mai multe exemple: http://en.wikipedia.org/wiki/Conjugate_prior_distribution.

2 Date continue cu a priori continue

2.1 Scopurile învățării

1. Să poată construi un tabel de actualizare Bayesiană pentru ipoteze continue și date continue.
2. Să poată recunoaște pdf a unei repartiții normale și să determine media și dispersia ei.

2.2 Introducere

Facem actualizare Bayesiană când atât ipotezele cât și datele iau valori continue. Modelul este același cu ce-am făcut înainte; vom recapitula mai întâi celelalte 2 cazuri.

2.3 Cazurile anterioare

2.3.1 Ipoteze discrete, date discrete

Notății

Ipoteze \mathcal{H}

Date x

A priori $P(\mathcal{H})$

Verosimilitatea $p(x|\mathcal{H})$

A posteriori $P(\mathcal{H}|x)$.

Exemplul 1. Presupunem că avem datele x și 3 posibile explicații (ipoteze) pentru date, pe care le vom numi A, B, C . De asemenea presupunem că datele pot lua 2 valori posibile, -1 și 1 .

Pentru a folosi datele pentru estimarea probabilităților diverselor ipoteze, avem nevoie de o pmf a priori și un tabel de verosimilitate. Presupunem că a priori și verosimilitățile sunt date în următoarele tabele.

| hypothesis \mathcal{H} | prior $P(\mathcal{H})$ |
|-----------------------------|---------------------------|
| A | 0.1 |
| B | 0.3 |
| C | 0.6 |

Probabilități a priori

| hypothesis \mathcal{H} | likelihood $p(x \mathcal{H})$ | |
|-----------------------------|-------------------------------|---------|
| | $x = -1$ | $x = 1$ |
| A | 0.2 | 0.8 |
| B | 0.5 | 0.5 |
| C | 0.7 | 0.3 |

Verosimilități

Fiecare element din tabelul de verosimilitate este o verosimilitate $p(x|\mathcal{H})$. De exemplu, $p(x = -1|A) = 0.2$.

Întrebare: Presupunem că obținem datele $x_1 = 1$. Folosiți aceasta pentru

a obține probabilitățile a posteriori pentru ipoteze.

Răspuns. Datele aleg o coloană din tabelul de verosimilități pe care o folosim apoi în tabelul nostru de actualizare Bayesiană.

| hypothesis | prior | likelihood | Bayes | |
|---------------|------------------|--------------------------|------------------------------------|--|
| | | | numerator | posterior |
| \mathcal{H} | $P(\mathcal{H})$ | $p(x = 1 \mathcal{H})$ | $p(x \mathcal{H})P(\mathcal{H})$ | $P(\mathcal{H} x) = \frac{p(x \mathcal{H})P(\mathcal{H})}{p(x)}$ |
| A | 0.1 | 0.8 | 0.08 | 0.195 |
| B | 0.3 | 0.5 | 0.15 | 0.366 |
| C | 0.6 | 0.3 | 0.18 | 0.439 |
| total | 1 | | $p(x) = 0.41$ | 1 |

Pentru a rezuma: probabilitățile a priori ale ipotezelor și verosimilitățile datelor cunoscând ipotezele au fost date; numărătorul Bayes este produsul dintre a priori și verosimilitate; probabilitatea totală $p(x)$ este suma probabilităților din coloana numărătorilor Bayes; împărțim la $p(x)$ pentru a normaliza numărătorii Bayes.

2.3.2 Ipoteze continue, date discrete

Acum presupunem că avem datele x care pot lua o mulțime discretă de valori și un parametru continuu θ care determină repartiția din care sunt extrase datele.

Notății

Ipoteze θ

Date x

A priori $f(\theta)d\theta$

Verosimilitate $p(x|\theta)$

A posteriori $f(\theta|x)d\theta$.

Notă: Am înmulțit cu $d\theta$ pentru a exprima a priori și a posteriori ca probabilități. Ca densități, avem pdf a priori $f(\theta)$ și pdf a posteriori $f(\theta|x)$.

Exemplul 2. Presupunem că $x \sim \text{Binomial}(5, \theta)$. Deci θ este în domeniul $[0, 1]$ și datele x pot lua 6 valori posibile, 0, 1, ..., 5.

Deoarece există un domeniu continuu de valori, folosim o pdf pentru a descrie a priori pentru θ . Presupunem că a priori este $f(\theta) = 2\theta$. Putem face totuși un tabel de verosimilitate, cu toate că are o singură linie, reprezentând o ipoteză arbitrară θ .

| hypothesis | likelihood $p(x \theta)$ | | | | | |
|------------|------------------------------|------------------------------------|--------------------------------------|--------------------------------------|------------------------------------|------------------------|
| | $x = 0$ | $x = 1$ | $x = 2$ | $x = 3$ | $x = 4$ | $x = 5$ |
| θ | $\binom{5}{0}(1 - \theta)^5$ | $\binom{5}{1}\theta(1 - \theta)^4$ | $\binom{5}{2}\theta^2(1 - \theta)^3$ | $\binom{5}{3}\theta^3(1 - \theta)^2$ | $\binom{5}{4}\theta^4(1 - \theta)$ | $\binom{5}{5}\theta^5$ |

Verosimilități

Întrebare Presupunem că obținem data $x_1 = 2$. Folosiți aceasta pentru a afla pdf a posteriori pentru parametrul (ipoteza) θ .

Răspuns. Ca mai înainte, data alege una din coloanele din tabelul de verosimilități pe care o putem folosi în tabelul nostru de actualizare Bayesiană. Deoarece vrem să lucrăm cu probabilități scriem $f(\theta)d\theta$ și $f(\theta|x_1)d\theta$. Pe ultima linie, în loc de " θ^2 " se va citi " θ^3 ". Pe ultima coloană, în loc de " $\frac{3!3!}{7!}$ " se va citi " $\frac{7!}{3!3!}$ ".

| hypothesis | prior | likelihood | Bayes numerator | posterior |
|------------|---------------------|--|---|---|
| θ | $f(\theta) d\theta$ | $p(x = 2 \theta)$ | $p(x \theta)f(\theta) d\theta$ | $f(\theta x) d\theta = \frac{p(x \theta)f(\theta) d\theta}{p(x)}$ |
| θ | $2\theta d\theta$ | $\binom{5}{2}\theta^2(1-\theta)^3$ | $2\binom{5}{2}\theta^3(1-\theta)^3 d\theta$ | $f(\theta x) d\theta = \frac{3!3!}{7!}\theta^3(1-\theta)^3 d\theta$ |
| total | 1 | $p(x) = \int_0^1 2\binom{5}{2}\theta^2(1-\theta)^3 d\theta = 2\binom{5}{2}\frac{3!3!}{7!}$ | | 1 |

Pentru a rezuma: probabilitățile a priori ale ipotezelor și verosimilitățile datelor cunoscând ipotezele sunt date; numărătorul Bayes este produsul dintre a priori și verosimilitate; probabilitatea totală $p(x)$ este integrala probabilităților din coloana numărătorului Bayes; împărțim prin $p(x)$ pentru a normaliza numărătorul Bayes.

2.4 Ipoteze continue și date continue

Când atât datele și ipotezele sunt continue, singura schimbare în exemplul anterior este că funcția de verosimilitate folosește o pdf $f(x|\theta)$ în locul unei pmf $p(x|\theta)$. Forma generală a tabelului de actualizare Bayesiană este aceeași.

Notății

Ipoteze θ

Date x

A priori $f(\theta)d\theta$

Verosimilitate $f(x|\theta)dx$

A posteriori $f(\theta|x)d\theta$.

Simplificarea notației. În cazurile precedente am inclus $d\theta$ astfel că lucrăm cu probabilități în loc de densități. Când atât datele cât și ipotezele sunt continue, vom avea nevoie atât de $d\theta$ cât și de dx . Aceasta face lucrurile mai simple conceptual, dar mai greoaie notațional. Pentru a simplifica notația ne vom permite să eliminăm dx din tabelele noastre. Aceasta este în regulă, deoarece datele x sunt fixate. Vom păstra $d\theta$ deoarece ipotezei θ i se permite să varieze.

Pentru comparație, întâi arătăm tabelul general cu notație simplificată, urmat imediat apoi de tabelul arătând infinezimalele. La primul tabel. pe ultima coloană, în loc de " $f(\theta|x)$ " se va citi " $f(\theta|x)d\theta$ ".

| hypoth. | prior | likelihood | Bayes numerator | posterior |
|----------|---------------------|--|--------------------------------|---|
| θ | $f(\theta) d\theta$ | $f(x \theta)$ | $f(x \theta)f(\theta) d\theta$ | $f(\theta x) = \frac{f(x \theta)f(\theta) d\theta}{f(x)}$ |
| total | 1 | $f(x) = \int f(x \theta)f(\theta) d\theta$ | | 1 |

Tabel de actualizare Bayesiană fără dx

| hypoth. | prior | likelihood | Bayes numerator | posterior |
|----------|---------------------|--|-----------------------------------|---|
| θ | $f(\theta) d\theta$ | $f(x \theta) dx$ | $f(x \theta)f(\theta) d\theta dx$ | $f(\theta x) d\theta = \frac{f(x \theta)f(\theta) d\theta dx}{f(x) dx} = \frac{f(x \theta)f(\theta) d\theta}{f(x)}$ |
| total | 1 | $f(x) dx = (\int f(x \theta)f(\theta) d\theta) dx$ | | 1 |

Tabel de actualizare Bayesiană cu $d\theta$ și dx

Pentru a rezuma: probabilitățile a priori ale ipotezelor și verosimilitățile datelor cunoscând ipotezele erau date; numărătorul Bayes este produsul dintre a priori și verosimilitate; probabilitatea totală $f(x)dx$ este integrala probabilităților din coloana numărătorului Bayes; împărțim prin $f(x)dx$ pentru a normaliza numărătorul Bayes.

2.5 Ipoteză normală, date normale

Un exemplu standard de ipoteze continue și date continue presupune că atât datele cât și a priori au repartiții normale. Următorul exemplu presupune că dispersia datelor este cunoscută.

Exemplul 3. Presupunem că avem data $x = 5$ care a fost extrasă dintr-o repartiție normală cu medie necunoscută θ și deviație standard 1.

$$x \sim N(\theta, 1).$$

Presupunem mai departe că repartiția noastră a priori pentru θ este $\theta \sim N(2, 1)$.

Fie x o valoare arbitrară a datelor.

- Faceți un tabel Bayesian cu a priori, verosimilitate și numărător Bayes.
- Arătați că repartiția a posteriori pentru θ este tot normală.

c) Aflați media și dispersia repartiției a posteriori.

Răspuns. Cum am făcut cu tabelele de mai sus, un bun compromis asupra notației este să includem $d\theta$, dar nu dx . Motivul pentru aceasta este că probabilitatea totală este calculată integrând după θ și $d\theta$ ne reamintește asta.

Pdf a priori a noastră este

$$f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2}.$$

Funcția de verosimilitate este

$$f(x=5|\theta) = \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2}.$$

Înmulțim a priori cu verosimilitatea.

$$\begin{aligned} \text{a priori} \cdot \text{verosimilitatea} &= \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2} \\ &= \frac{1}{2\pi} e^{-(2\theta^2 - 14\theta + 29)/2} \\ &= \frac{1}{2\pi} e^{-(\theta^2 - 7\theta + 29/2)} \quad (\text{completăm pătratul}) \\ &= \frac{1}{2\pi} e^{-((\theta-7/2)^2 + 9/4)} \\ &= \frac{e^{-9/4}}{2\pi} e^{-(\theta-7/2)^2} \\ &= c_1 e^{-(\theta-7/2)^2}. \end{aligned}$$

La ultimul pas am înlocuit factorul constant complicat cu expresia mai simplă c_1 . Pe linia a 3-a a tabelului, în loc de " $e^{-(\theta-7/2)^2}$ " se va citi " $e^{-(\theta-7/2)^2} d\theta$ " (de 2 ori).

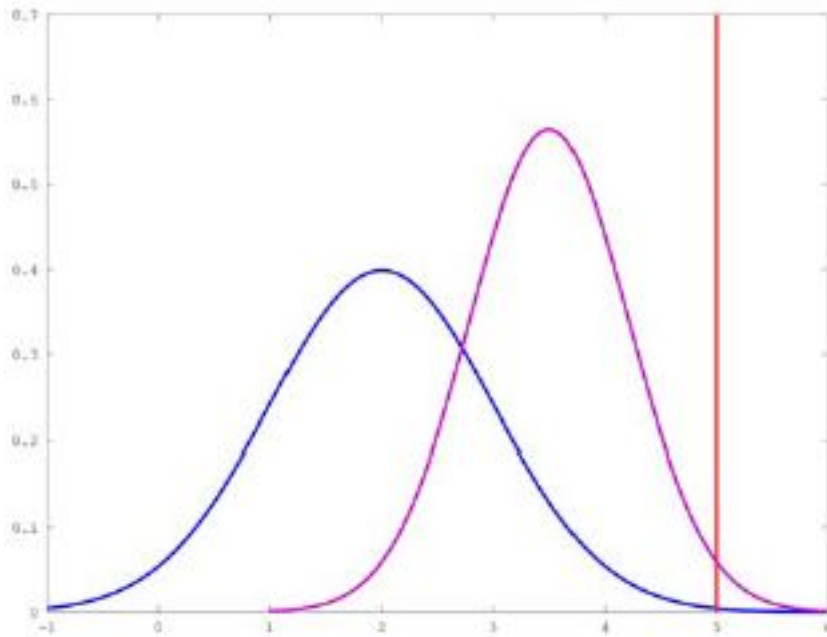
| hypothesis | prior | likelihood | Bayes numerator | posterior $f(\theta x=5) d\theta$ |
|------------|---|---|--|---|
| θ | $f(\theta) d\theta$ | $f(x=5 \theta)$ | $f(x=5 \theta)f(\theta) d\theta$ | $\frac{f(x=5 \theta)f(\theta) d\theta}{f(x=5)}$ |
| θ | $\frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} d\theta$ | $\frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2}$ | $c_1 e^{-(\theta-7/2)^2}$ | $c_2 e^{-(\theta-7/2)^2}$ |
| total | 1 | | $f(x=5) = \int f(x=5 \theta)f(\theta) d\theta$ | 1 |

Pdf a posteriori este cea a unei repartiții normale. Deoarece exponențiala unei repartiții normale este $e^{-(\theta-\mu)^2/2\sigma^2}$, avem media $\mu = 7/2$ și $2\sigma^2 = 1$, deci

dispersia este $\sigma^2 = 1/2$.

Nu trebuie să calculăm probabilitatea totală; ea este folosită doar pentru normalizare și știm deja constanta de normalizare $\frac{1}{\sigma\sqrt{2\pi}}$ pentru o repartiție normală.

Iată graficele pdf-urilor a priori și a posteriori. Observăm cum data "trage" a priori spre ea.



a priori = albastru; a posteriori = mov; data = roșu

Acum, repetăm exemplul precedent pentru x general.

Exemplul 4. Presupunem că data noastră x este extrasă dintr-o repartiție normală cu medie necunoscută θ și deviație standard 1.

$$x \sim N(\theta, 1).$$

Răspuns. Pdf a priori și funcția de verosimilitate sunt

$$f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2}, \quad f(x|\theta) = \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}.$$

Numărătorul Bayes este produsul dintre a priori și verosimilitate:

$$\begin{aligned}
 \text{a priori} \cdot \text{verosimilitatea} &= \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2} \\
 &= \frac{1}{2\pi} e^{-(2\theta^2 - (4+2x)\theta + 4+x^2)/2} \\
 &= \frac{1}{2\pi} e^{-(\theta^2 - (2+x)\theta + (4+x^2)/2)} \quad (\text{completăm pătratul}) \\
 &= \frac{1}{2\pi} e^{-((\theta - (1+x/2))^2 - (1+x/2)^2 + (4+x^2)/2)} \\
 &= c_1 e^{-(\theta - (1+x/2))^2}.
 \end{aligned}$$

Ca în ultimul exemplu, în ultimul pas am înlocuit toate constantele, inclusiv exponențialele care implică doar pe x , prin simpla constantă c_1 .

Acum, tabelul Bayesian de înlocuire devine (pe linia a 3-a a tabelului, în loc de " $e^{-(\theta - (1+x/2))^2}$ " se va citi " $e^{-(\theta - (1+x/2))^2} d\theta$ " (de 2 ori); pe ultima linie a tabelului, în loc de "=" se va citi " $= \int$ ").

| hypothesis | prior | likelihood | Bayes numerator | posterior $f(\theta x) d\theta$ |
|------------|---|---|-----------------------------------|--|
| θ | $f(\theta) d\theta$ | $f(x \theta)$ | $f(x \theta) f(\theta) d\theta$ | $\frac{f(x \theta) f(\theta) d\theta}{f(x)}$ |
| θ | $\frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} d\theta$ | $\frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}$ | $c_1 e^{-(\theta - (1+x/2))^2}$ | $c_2 e^{-(\theta - (1+x/2))^2}$ |
| total | 1 | $f(x) = \int f(x \theta) f(\theta) d\theta$ | | 1 |

Ca în exemplul precedent putem vedea din forma a posteriori că ea trebuie să fie a unei repartiții normale cu media $1 + x/2$ și dispersia $1/2$. (Comparați aceasta cu cazul $x = 5$ din exemplul precedent.)

2.6 Probabilități predictive

Deoarece datele x sunt continue, au pdf-uri predictive a priori și a posteriori. Pdf predictivă a priori este densitatea de probabilitate totală calculată la marginea de jos a coloanei numărătorului Bayes:

$$f(x) = \int f(x|\theta) f(\theta) d\theta,$$

unde integrala este calculată pe întregul domeniu al lui θ .

Pdf predictivă a posteriori are aceeași formă ca pdf predictivă a priori, cu

excepția faptului că folosește probabilitățile a posteriori pentru θ :

$$f(x_2|x_1) = \int f(x_2|\theta, x_1)f(\theta|x_1)d\theta.$$

Ca de obicei, presupunem că x_1 și x_2 sunt **condiționat independente**. Adică,

$$f(x_2|\theta, x_1) = f(x_2|\theta).$$

În acest caz formula pentru pdf predictivă a posteriori este un pic mai simplă:

$$f(x_2|x_1) = \int f(x_2|\theta)f(\theta|x_1)d\theta.$$

3 A priori conjugate: Beta și normală

3.1 Scopurile învățării

1. Să înțeleagă beneficiile a priori conjugate.
2. Să poată să actualizeze o a priori beta dată fiind o verosimilitate Bernoulli, binomială sau geometrică.
3. Să înțeleagă și să poată folosi formula pentru actualizarea unei a priori normale fiind dată o verosimilitate normală cu dispersie cunoscută.

3.2 Introducere și definiție

Cu o a priori conjugată, a posteriori este de același tip, de exemplu pentru verosimilitate binomială, a priori beta devine o a posteriori beta. A priori conjugate sunt utile deoarece ele reduc actualizarea Bayesiană la modificarea parametrilor repartiției a priori (așa numiții hiperparametri) în locul calculului de integrale.

Ne vom concentra pe 2 exemple importante de a priori conjugate: beta și normală. O listă mult mai cuprinzătoare este în tabelele din http://en.wikipedia.org/wiki/Conjugate_prior_distribution.

Definiție. Presupunem că avem date cu funcția de verosimilitate $f(x|\theta)$ depinzând de un parametru θ . Mai presupunem că repartiția a priori pentru θ este una dintr-o familie de repartiții parametrizate. Dacă repartiția a posteriori pentru θ este în aceeași familie ca repartiția a priori, spunem că a priori este o **a priori conjugată** pentru verosimilitate.

3.3 Repartiția beta

Arătăm că repartiția beta este o a priori conjugată pentru verosimilități binomiale, Bernoulli și geometrice.

3.3.1 Verosimilitate binomială

Am văzut că [repartiția beta este o a priori conjugată pentru repartiția binomială](#). Aceasta înseamnă că dacă funcția de verosimilitate este binomială și repartiția a priori este beta, atunci repartiția a posteriori este tot beta.

Mai concret, presupunem că funcția de verosimilitate are o repartiție binomială(N, θ), unde N este cunoscut și θ este parametrul (necunoscut) de interes. Avem de asemenea că data x este un întreg între 0 și N . Atunci, pentru o a priori beta avem următorul tabel:

| hypothesis | data | prior | likelihood | posterior |
|------------|------|---------------------------------------|-----------------------------------|---|
| θ | x | $\text{beta}(a, b)$ | $\text{binomial}(N, \theta)$ | $\text{beta}(a + x, b + N - x)$ |
| θ | x | $c_1 \theta^{a-1} (1 - \theta)^{b-1}$ | $c_2 \theta^x (1 - \theta)^{N-x}$ | $c_3 \theta^{a+x-1} (1 - \theta)^{b+N-x-1}$ |

Tabelul este simplificat scriind coeficienții de normalizare ca c_1, c_2 și c_3 .

$$c_1 = \frac{(a + b - 1)!}{(a - 1)!(b - 1)!}, \quad c_2 = C_N^x = \frac{N!}{x!(N - x)!}, \quad c_3 = \frac{(a + b + N - 1)!}{(a + x - 1)!(b + N - x - 1)!}.$$

3.3.2 Verosimilitate Bernoulli

[Repartiția beta este o a priori conjugată pentru repartiția Bernoulli](#). Acesta este de fapt un caz special al repartiției binomiale, deoarece $\text{Bernoulli}(\theta)$ este aceeași ca $\text{binomial}(1, \theta)$. În tabelul de mai jos, arătăm actualizările corespunzând succesului ($x = 1$) și eșecului ($x = 0$) pe linii separate.

| hypothesis | data | prior | likelihood | posterior |
|------------|---------|---------------------------------------|----------------------------|--|
| θ | x | $\text{beta}(a, b)$ | $\text{Bernoulli}(\theta)$ | $\text{beta}(a + 1, b)$ or $\text{beta}(a, b + 1)$ |
| θ | $x = 1$ | $c_1 \theta^{a-1} (1 - \theta)^{b-1}$ | θ | $c_3 \theta^a (1 - \theta)^{b-1}$ |
| θ | $x = 0$ | $c_1 \theta^{a-1} (1 - \theta)^{b-1}$ | $1 - \theta$ | $c_3 \theta^{a-1} (1 - \theta)^b$ |

Constantele c_1 și c_3 au aceleași formule ca în cazul precedent (al verosimilității binomiale) cu $N = 1$.

3.3.3 Verosimilitate geometrică

Repartiția geometrică(θ) descrie probabilitatea a x eșecuri înaintea primului succes, unde probabilitatea succesului în fiecare încercare independentă este θ . Pmf corespunzătoare este $p(x) = \theta(1 - \theta)^x$.

Acum presupunem că avem o dată x și ipoteza noastră θ este că x este extrasă dintr-o repartiție geometrică(θ). Din tabel vedem că [repartiția beta este o a priori conjugată pentru o verosimilitate geometrică](#):

| ipoteza | data | a priori | verosimilitatea | a posteriori |
|----------|------|---------------------------------------|-----------------------------|-------------------------------------|
| θ | x | $\text{beta}(a, b)$ | $\text{geometrică}(\theta)$ | $\text{beta}(a + 1, b + x)$ |
| θ | x | $c_1 \theta^{a-1} (1 - \theta)^{b-1}$ | $\theta(1 - \theta)^x$ | $c_3 \theta^a (1 - \theta)^{b+x-1}$ |

Exemplul 1. În timp ce călătoreau prin Regatul Ciupercilor, Mario și Luigi

au găsit niște monede neobișnuite. Ei au căzut de acord asupra unei a priori $f(\theta) \sim \text{beta}(5, 5)$ pentru probabilitatea aversului, dar nu au fost de acord ce experiment să facă pentru a investiga θ .

a) Mario decide să arunce o monedă de 5 ori. El obține un avers în 5 aruncări.

b) Luigi decide să arunce o monedă până la primul avers. El obține 4 reversuri înaintea primului avers.

Arătați că Mario și Luigi ajung la aceeași a posteriori pentru θ și calculați această a posteriori.

Răspuns. Tabelul lui Mario:

| ipoteza | data | a priori | verosimilitatea | a posteriori |
|----------|---------|-----------------------------|-----------------------------|-----------------------------|
| θ | $x = 1$ | $\text{beta}(5, 5)$ | $\text{binomială}(\theta)$ | ??? |
| θ | $x = 1$ | $c_1\theta^4(1 - \theta)^4$ | $C_5^1\theta(1 - \theta)^4$ | $c_3\theta^5(1 - \theta)^8$ |

Tabelul lui Luigi:

| ipoteza | data | a priori | verosimilitatea | a posteriori |
|----------|---------|-----------------------------|-----------------------------|-----------------------------|
| θ | $x = 4$ | $\text{beta}(5, 5)$ | $\text{geometrică}(\theta)$ | ??? |
| θ | $x = 4$ | $c_1\theta^4(1 - \theta)^4$ | $\theta(1 - \theta)^4$ | $c_3\theta^5(1 - \theta)^8$ |

Atât a posteriori a lui Mario cât și a lui Luigi au forma unei repartiții $\text{beta}(6, 9)$. Factorul de normalizare este același în ambele cazuri deoarece este determinat cerând ca probabilitatea totală să fie 1.

3.4 Normala generează normală

Repartiția normală este a priori conjugată cu ea însăși. În particular, dacă funcția de verosimilitate este normală cu dispersie cunoscută, atunci o a priori normală dă o a posteriori normală. Acum atât ipotezele și datele sunt continue.

Presupunem că avem o măsurare $x \sim N(\theta, \sigma^2)$, unde dispersia σ^2 este cunoscută. Adică, media θ este parametrul nostru necunoscut de interes și știm că verosimilitatea vine dintr-o repartiție normală cu dispersia σ^2 . Dacă alegem o pdf a priori normală

$$f(\theta) \sim N(\mu_{\text{prior}}, \sigma_{\text{prior}}^2),$$

atunci pdf a posteriori este de asemenea normală: $f(\theta|x) \sim N(\mu_{\text{post}}, \sigma_{\text{post}}^2)$, unde

$$\frac{\mu_{\text{post}}}{\sigma_{\text{post}}^2} = \frac{\mu_{\text{prior}}}{\sigma_{\text{prior}}^2} + \frac{x}{\sigma^2}, \quad \frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_{\text{prior}}^2} + \frac{1}{\sigma^2}. \quad (1)$$

Următoarea formă a acestor formule este mai ușor de citit și arată că μ_{post} este o medie ponderată între μ_{prior} și data x .

$$a = \frac{1}{\sigma_{\text{prior}}^2}, \quad b = \frac{1}{\sigma^2}, \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + bx}{a + b}, \quad \sigma_{\text{post}}^2 = \frac{1}{a + b}. \quad (2)$$

Cu aceste formule în minte, putem exprima actualizarea prin tabelul:

| hypothesis | data | prior | likelihood | posterior |
|------------|------|--|--|--|
| θ | x | $f(\theta) \sim N(\mu_{\text{prior}}, \sigma_{\text{prior}}^2)$ | $f(x \theta) \sim N(\theta, \sigma^2)$ | $f(\theta x) \sim N(\mu_{\text{post}}, \sigma_{\text{post}}^2)$ |
| θ | x | $c_1 \exp\left(\frac{-(\theta - \mu_{\text{prior}})^2}{2\sigma_{\text{prior}}^2}\right)$ | $c_2 \exp\left(\frac{-(x - \theta)^2}{2\sigma^2}\right)$ | $c_3 \exp\left(\frac{-(\theta - \mu_{\text{post}})^2}{2\sigma_{\text{post}}^2}\right)$ |

Demonstrația formulelor generale se face analog ca în următorul exemplu numeric.

Exemplul 2. Presupunem că avem a priori $\theta \sim N(4, 8)$ și verosimilitatea $x \sim N(\theta, 5)$. Presupunem de asemenea că avem o măsurare $x_1 = 3$. Arătați că repartiția a posteriori este normală.

Răspuns.

a priori: $f(\theta) = c_1 e^{-(\theta-4)^2/16}$; verosimilitatea: $f(x_1|\theta) = c_2 e^{-(x_1-\theta)^2/10} = c_2 e^{-(3-\theta)^2/10}$.

Înmulțim a priori cu verosimilitatea pentru a obține a posteriori:

$$f(\theta|x_1) = c_1 c_2 e^{-(\theta-4)^2/16} e^{-(3-\theta)^2/10} = c_1 c_2 \exp\left(-\frac{(\theta-4)^2}{16} - \frac{(3-\theta)^2}{10}\right).$$

Completăm pătratul din exponent

$$\begin{aligned} -\frac{(\theta-4)^2}{16} - \frac{(3-\theta)^2}{10} &= -\frac{5(\theta-4)^2 + 8(3-\theta)^2}{80} \\ &= -\frac{13\theta^2 - 88\theta + 152}{80} \\ &= -\frac{\theta^2 - \frac{88}{13}\theta + \frac{152}{13}}{80/13} \\ &= -\frac{(\theta - 44/13)^2 + 152/13 - (44/13)^2}{80/13}. \end{aligned}$$

De aceea, a posteriori este

$$f(\theta|x_1) = c_1 c_2 e^{-\frac{(\theta-44/13)^2 + 152/13 - (44/13)^2}{80/13}} = c_3 e^{-\frac{(\theta-44/13)^2}{80/13}}.$$

Aceasta are forma pdf pentru $N(44/13, 40/13)$, q.e.d.

Verificăm aceasta cu formulele (2).

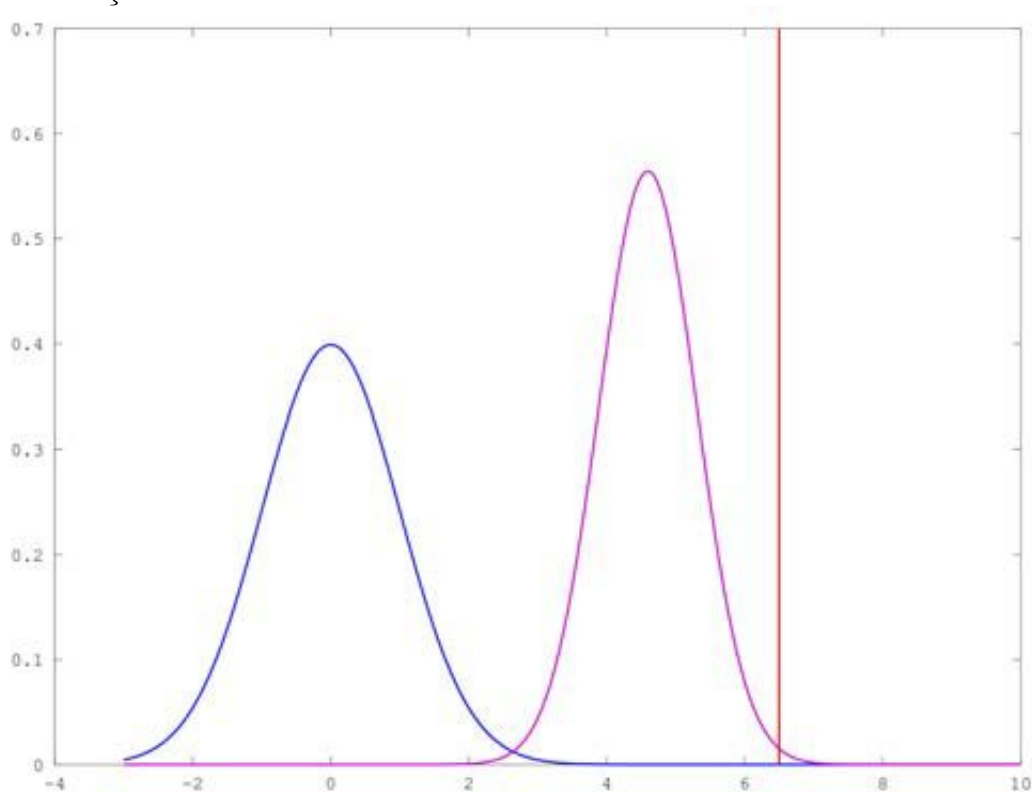
$$\mu_{\text{prior}} = 4, \sigma_{\text{prior}}^2 = 8, \sigma^2 = 5 \implies a = \frac{1}{8}, b = \frac{1}{5}.$$

De aceea

$$\begin{aligned} \mu_{\text{post}} &= \frac{a\mu_{\text{prior}} + bx}{a+b} = \frac{\frac{1}{8} \cdot 4 + \frac{1}{5} \cdot 3}{\frac{1}{8} + \frac{1}{5}} = \frac{44}{13} \approx 3.38, \\ \sigma_{\text{post}}^2 &= \frac{1}{a+b} = \frac{1}{\frac{1}{8} + \frac{1}{5}} = \frac{40}{13} \approx 3.08. \end{aligned}$$

Exemplul 3. Presupunem că știm datele $x \sim N(\theta, \sigma^2)$ și avem a priori $N(0, 1)$. Obținem o valoare a datelor $x = 6.5$. Descrieți schimbările pdf pentru θ în actualizarea de la a priori la a posteriori.

Răspuns. Iată graficul pdf-urilor a priori, a posteriori cu data marcată cu o line roșie.



A priori în albastru, a posteriori în mov, data în roșu.

Media a posteriori va fi o medie ponderată dintre media a priori și dată. Vârful pdf a posteriori va fi între vârful a priori și linia roșie. Avem

$$\sigma_{\text{post}}^2 = \frac{1}{1/\sigma_{\text{prior}}^2 + 1/\sigma^2} = \sigma_{\text{prior}}^2 \cdot \frac{\sigma^2}{\sigma_{\text{prior}}^2 + \sigma^2} < \sigma_{\text{prior}}^2.$$

Adică a posteriori are dispersie mai mică decât a priori, i.e. data ne face mai siguri despre unde este θ în domeniul său.

3.4.1 Mai mult de o dată

Exemplul 4. Presupunem că avem datele x_1, x_2, x_3 . Folosiți formulele (1) pentru a actualiza succesiv.

Răspuns. Notăm media și dispersia a priori cu μ_0 , respectiv σ_0^2 . Mediile și

dispersiile actualizate vor fi μ_i , respectiv σ_i . Avem succesiv

$$\begin{aligned}\frac{1}{\sigma_1^2} &= \frac{1}{\sigma_0^2} + \frac{1}{\sigma^2}; \quad \frac{\mu_1}{\sigma_1^2} = \frac{\mu_0}{\sigma_0^2} + \frac{x_1}{\sigma^2} \\ \frac{1}{\sigma_2^2} &= \frac{1}{\sigma_1^2} + \frac{1}{\sigma^2} = \frac{1}{\sigma_0^2} + \frac{2}{\sigma^2}; \quad \frac{\mu_2}{\sigma_2^2} = \frac{\mu_1}{\sigma_1^2} + \frac{x_2}{\sigma^2} = \frac{\mu_0}{\sigma_0^2} + \frac{x_1 + x_2}{\sigma^2} \\ \frac{1}{\sigma_3^2} &= \frac{1}{\sigma_2^2} + \frac{1}{\sigma^2} = \frac{1}{\sigma_0^2} + \frac{3}{\sigma^2}; \quad \frac{\mu_3}{\sigma_3^2} = \frac{\mu_2}{\sigma_2^2} + \frac{x_3}{\sigma^2} = \frac{\mu_0}{\sigma_0^2} + \frac{x_1 + x_2 + x_3}{\sigma^2}.\end{aligned}$$

Exemplul se generalizează la n valori ale datelor x_1, \dots, x_n :

Formule de actualizare normală-normală pentru n date

$$\frac{\mu_{\text{post}}}{\sigma_{\text{post}}^2} = \frac{\mu_{\text{prior}}}{\sigma_{\text{prior}}^2} + \frac{n\bar{x}}{\sigma^2}, \quad \frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_{\text{prior}}^2} + \frac{n}{\sigma^2}, \quad \bar{x} = \frac{x_1 + \dots + x_n}{n}. \quad (3)$$

Din nou dăm o formă mai simplă de citit, arătând că μ_{post} este o medie ponderată între μ_{prior} și media de selecție \bar{x} :

$$a = \frac{1}{\sigma_{\text{prior}}^2}, \quad b = \frac{n}{\sigma^2}, \quad \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\bar{x}}{a + b}, \quad \sigma_{\text{post}}^2 = \frac{1}{a + b}. \quad (4)$$

Interpretare: μ_{post} este o medie ponderată între μ_{prior} și \bar{x} . Dacă numărul datelor este mare, atunci ponderea b este mare și \bar{x} va avea o puternică influență asupra a posteriori. Dacă σ_{prior}^2 este mică, atunci ponderea a este mare și μ_{prior} va avea o puternică influență asupra a posteriori. Pentru a rezuma:

1. Multe date au o mare influență asupra a posteriori.
2. Siguranța mare (dispersia mică) în a priori are o mare influență asupra a posteriori.