Curs 9

Cristian Niculescu

1 Repartiţii beta

1.1 Scopurile învăţării

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

1.2 Repartiţia beta

Repartiţia beta beta(a, b) este o repartiţie cu 2 parametri cu domeniul $[0, 1]$ şi pdf

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

Următoarea aplicaţie ne permite să explorăm forma repartiţiei Beta când parametrii variază:
`http://mathlets.org/mathlets/beta-distribution/`.
Repartiţia beta poate fi definită pentru orice numere reale $a > 0$ şi $b > 0$. Am definit-o doar pentru $a, b \in \mathbb{N}^*$, dar puteţi vedea întreaga istorie aici: `http://en.wikipedia.org/wiki/Beta_distribution`.
În contextul actualizării bayesiene, a şi b sunt numiţi adesea hiperparametri pentru a-i deosebi de parametrul necunoscut θ care reprezintă ipoteza noastră. Într-un anumit sens, a şi b sunt la "un nivel mai sus" decât θ, deoarece ei parametrizează pdf.

1.2.1 O observaţie 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 repartiţii beta(a, b) şi 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 şi ea este dată în formula pentru repartiţia beta.
O observaţie similară este valabilă pentru repartiţiile normală, exponenţială, etc.

1.2.2 A priori şi 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 şi obţinem 8 aversuri şi 4 reversuri. Plecând cu o a priori plată, arătaţi că pdf a posteriori este o repartiţie 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. Observaţia noastră simplă ne va spune că acesta trebuie să fie factorul constant din pdf beta.
Datele sunt 8 aversuri şi 4 reversuri. Deoarece acestea vin dintr-o repartiţie binomială$(12, \theta)$, 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

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

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

are o repartiţie beta$(9, 5)$ şi 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 şi să nu-i dăm valoarea explicită.
Exemplul 2. Acum presupunem că aruncăm aceeaşi monedă din nou, obţinând n aversuri şi m reversuri. Folosind pdf a posteriori din exemplul precedent ca noua noastră pdf a priori, arătaţi că noua pdf a posteriori este cea a unei repartiţii beta$(9 + n, 5 + m)$.
Răspuns. Totul este în tabel. Vom numi x_2 datele acestor $n + m$ aruncări adiţionale. 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	$c_2\theta^8(1-\theta)^4\,d\theta$	$c_3\,\theta^n(1-\theta)^m$	$c_2c_3\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$	$c_4\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$
total	1		$T=\int_0^1 c_2c_3\,\theta^{n+8}(1-\theta)^{m+4}\,d\theta$	1

Din nou observaţia noastră simplă are loc şi, de aceea, pdf a posteriori

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

este cea a unei repartiţii beta$(n+9,m+5)$.
Observaţie. Beta plată. Repartiţia beta$(1,1)$ coincide cu repartiţia uniformă pe $[0,1]$, pe care am numit-o de asemenea a priori plată pentru θ. Aceasta rezultă înlocuind $a=1$ şi $b=1$ în definiţia repartiţiei beta, dând $f(\theta)=1$.
Rezumat. Dacă probabilitatea aversului este θ, numărul de aversuri în $n+m$ aruncări are o repartiţie binomială$(n+m,\theta)$. Am văzut că dacă a priori pentru θ este o repartiţie beta, atunci la fel este şi a posteriori; doar parametrii a şi b ai repartiţiei 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

Repartiţia beta este numită o a priori conjugată pentru repartiţia binomială. Aceasta înseamnă că, dacă funcţia de verosimiliate este binomială, atunci o a priori beta dă o a posteriori beta. De fapt, repartiţia beta este o a priori conjugată şi pentru repartiţiile Bernoulli şi geometrică.
Un alt exemplu important: repartiţia normală îşi este propria a priori conjugată. În particular, dacă funcţia de verosimiliate 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 repartiţiei a priori (aşa numiţii hiperparametri) în locul calculului integralelor. Am văzut asta pentru repartiţia 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 învăţării

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

2.2 Introducere

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

2.3 Cazurile anterioare

2.3.1 Ipoteze discrete, date discrete

Notaţii
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 şi 3 posibile explicaţii (ipoteze) pentru date, pe care le vom numi A, B, C. De asemenea presupunem că datele pot lua 2 valori posibile, -1 şi 1.
Pentru a folosi datele pentru estimarea probabilităţilor diverselor ipoteze, avem nevoie de o pmf a priori şi un tabel de verosimilitate. Presupunem că a priori şi verosimilităţile sunt date în următoarele tabele.

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

Probabilităţi a priori

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

Verosimilităţi

Fiecare element din tabelul de verosimilitate este o verosimilitate $p(x|\mathcal{H})$. De exemplu, $p(x = -1|A) = 0.2$.
Întrebare: Presupunem că obţinem datele $x_1 = 1$. Folosiţi aceasta pentru

a obţine probabilităţile a posteriori pentru ipoteze.
Răspuns. Datele aleg o coloană din tabelul de verosimilităţi pe care o folosim apoi în tabelul nostru de actualizare Bayesiană.

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$p(x=1\,\vert\,\mathcal{H})$	$p(x\,\vert\,\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\,\vert\,x) = \dfrac{p(x\,\vert\,\mathcal{H})P(\mathcal{H})}{p(x)}$
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: probabilităţile a priori ale ipotezelor şi verosimilităţile datelor cunoscând ipotezele au fost date; numărătorul Bayes este produsul dintre a priori şi verosimilitate; probabilitatea totală $p(x)$ este suma probabilităţilor din coloana numărătorilor Bayes; împărţim 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 mulţime discretă de valori şi un parametru continuu θ care determină repartiţia din care sunt extrase datele.
Notaţii
Ipoteze θ
Date x
A priori $f(\theta)d\theta$
Verosimilitate $p(x|\theta)$
A posteriori $f(\theta|x)d\theta$.
Notă: Am înmulţit cu $d\theta$ pentru a exprima a priori şi a posteriori ca probabilităţi. Ca densităţi, avem pdf a priori $f(\theta)$ şi pdf a posteriori $f(\theta|x)$.
Exemplul 2. Presupunem că $x \sim \text{Binomial}(5, \theta)$. Deci θ este în domeniul $[0, 1]$ şi 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 totuşi un tabel de verosimilitate, cu toate că are o singură linie, reprezentând o ipoteză arbitrară θ.

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

Verosimilităţi
Întrebare Presupunem că obţinem data $x_1 = 2$. Folosiţi aceasta pentru a afla pdf a posteriori pentru parametrul (ipoteza) θ.
Răspuns. Ca mai înainte, data alege una din coloanele din tabelul de verosimilităţi pe care o putem folosi în tabelul nostru de actualizare Bayesiană. Deoarece vrem să lucrăm cu probabilităţi scriem $f(\theta)d\theta$ şi $f(\theta|x_1)d\theta$. 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: probabilităţile a priori ale ipotezelor şi verosimilităţile datelor cunoscând ipotezele sunt date; numărătorul Bayes este produsul dintre a priori şi verosimilitate; probabilitatea totală $p(x)$ este integrala probabilităţilor din coloana numărătorului Bayes; împărţim prin $p(x)$ pentru a normaliza numărătorul Bayes.

2.4 Ipoteze continue şi date continue

Când atât datele şi ipotezele sunt continue, singura schimbare în exemplul anterior este că funcţia de verosimilitate foloseşte o pdf $f(x|\theta)$ în locul unei pmf $p(x|\theta)$. Forma generală a tabelului de actualizare Bayesiană este aceeaşi.
Notaţii
Ipoteze θ
Date x
A priori $f(\theta)d\theta$
Verosimilitate $f(x|\theta)dx$
A posteriori $f(\theta|x)d\theta$.
Simplificarea notaţiei. În cazurile precedente am inclus $d\theta$ astfel că lucram cu probabilităţi în loc de densităţi. Când atât datele cât şi ipotezele sunt continue, vom avea nevoie atât de $d\theta$ cât şi de dx. Aceasta face lucrurile mai simple conceptual, dar mai greoaie notaţional. Pentru a simplifica notaţia 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 comparaţie, întâi arătăm tabelul general cu notaţie simplificată, urmat imediat apoi de tabelul arătând infinitezimalele. 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	$f(\theta)\,d\theta$	$f(x\,\|\,\theta)$	$f(x\,\|\,\theta)f(\theta)\,d\theta$	$f(\theta\,\|\,x) = \dfrac{f(x\,\|\,\theta)f(\theta)\,d\theta}{f(x)}$
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	$f(\theta)\,d\theta$	$f(x\,\|\,\theta)\,dx$	$f(x\,\|\,\theta)f(\theta)\,d\theta\,dx$	$f(\theta\,\|\,x)\,d\theta = \dfrac{f(x\,\|\,\theta)f(\theta)\,d\theta\,dx}{f(x)\,dx} = \dfrac{f(x\,\|\,\theta)f(\theta)\,d\theta}{f(x)}$
total	1		$f(x)\,dx = \left(\int f(x\,\|\,\theta)f(\theta)\,d\theta\right)\,dx$	1

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

Pentru a rezuma: probabilităţile a priori ale ipotezelor şi verosimilităţile datelor cunoscând ipotezele erau date; numărătorul Bayes este produsul dintre a priori şi verosimilitate; probabilitatea totală $f(x)dx$ este integrala probabilităţilor din coloana numărătorului Bayes; împărţim prin $f(x)dx$ pentru a normaliza numărătorul Bayes.

2.5 Ipoteză normală, date normale

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

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

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

Presupunem mai departe că repartiţia noastră a priori pentru θ este
$\theta \sim N(2, 1)$.
Fie x o valoare arbitrară a datelor.
a) Faceţi un tabel Bayesian cu a priori, verosimilitate şi numărător Bayes.
b) Arătaţi că repartiţia a posteriori pentru θ este tot normală.

c) Aflaţi media şi dispersia repartiţiei a posteriori.
Răspuns. Cum am făcut cu tabelele de mai sus, un bun compromis asupra notaţiei este să includem $d\theta$, dar nu dx. Motivul pentru aceasta este că probabilitatea totală este calculată integrând după θ şi $d\theta$ ne reaminteşte asta.
Pdf a priori a noastră este

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

Funcţia de verosimilitate este

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

Înmulţim a priori cu verosimilitatea.

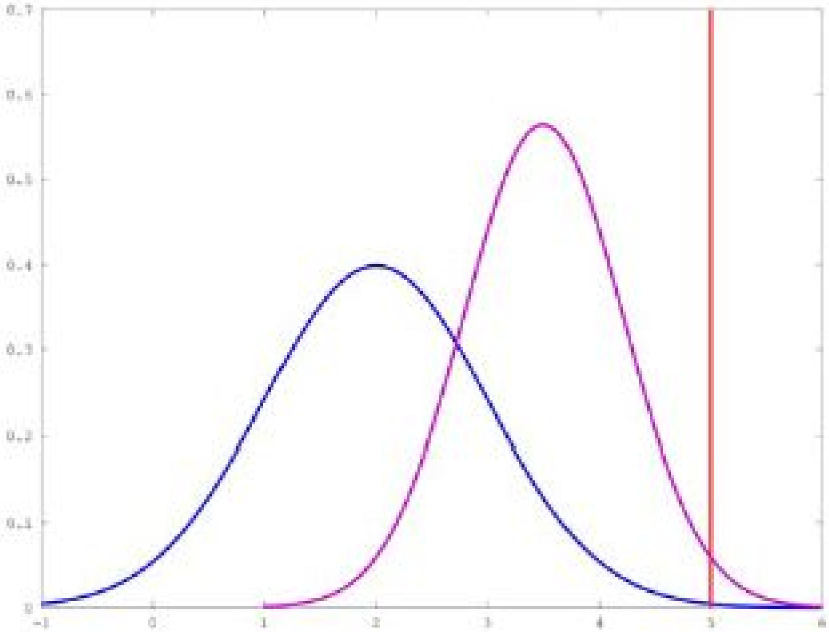
$$\begin{aligned}
\text{a priori} \cdot \text{verosimilitatea} &= \frac{1}{\sqrt{2\pi}} e^{-(\theta-2)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(5-\theta)^2/2} \\
&= \frac{1}{2\pi} e^{-(2\theta^2-14\theta+29)/2} \\
&= \frac{1}{2\pi} e^{-(\theta^2-7\theta+29/2)} \text{ (completăm pătratul)} \\
&= \frac{1}{2\pi}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

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 $f(\theta \mid x = 5)\, d\theta$
θ	$f(\theta)\, d\theta$	$f(x = 5 \mid \theta)$	$f(x = 5 \mid \theta) f(\theta)\, d\theta$	$\frac{f(x = 5 \mid \theta) f(\theta)\, d\theta}{f(x = 5)}$
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		$f(x = 5) = \int f(x = 5 \mid \theta) f(\theta)\, d\theta$	1

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

dispersia este $\sigma^2 = 1/2$.
Nu trebuie să calculăm probabilitatea totală; ea este folosită doar pentru normalizare şi ştim deja constanta de normalizare $\frac{1}{\sigma\sqrt{2\pi}}$ pentru o repartiţie normală.
Iată graficele pdf-urilor a priori şi a posteriori. Observăm cum data "trage" a priori spre ea.



a priori = albastru; a posteriori = mov; data = roşu
Acum, repetăm exemplul precedent pentru x general.
Exemplul 4. Presupunem că data noastră x este extrasă dintr-o repartiţie normală cu medie necunoscută θ şi deviaţie standard 1.

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

Răspuns. Pdf a priori şi funcţia de verosimilitate sunt

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

Numărătorul Bayes este produsul dintre a priori şi verosimilitate:

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

Ca în ultimul exemplu, în ultimul pas am înlocuit toate constantele, inclusiv exponenţialele 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)).

hypothesis	prior	likelihood	Bayes numerator	posterior $f(\theta\,\vert\,x)\,d\theta$
θ	$f(\theta)\,d\theta$	$f(x\,\vert\,\theta)$	$f(x\,\vert\,\theta)f(\theta)\,d\theta$	$\frac{f(x\,\vert\,\theta)f(\theta)\,d\theta}{f(x)}$
θ	$\frac{1}{\sqrt{2\pi}}e^{-(\theta-2)^2/2}\,d\theta$	$\frac{1}{\sqrt{2\pi}}e^{-(x-\theta)^2/2}$	$c_1e^{-(\theta-(1+x/2))^2}$	$c_2e^{-(\theta-(1+x/2))^2}$
total	1		$f(x) =$ $f(x\,\vert\,\theta)f(\theta)\,d\theta$	1

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

2.6 Probabilităţi predictive

Deoarece datele x sunt continue, au pdf-uri predictive a priori şi 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 aceeaşi formă ca pdf predictivă a priori, cu excepţia faptului că foloseşte probabilităţile 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 şi x_2 sunt condiţionat 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 şi normală

3.1 Scopurile învăţării

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

3.2 Introducere şi definiţie

Cu o a priori conjugată, a posteriori este de acelaşi 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 repartiţiei a priori (aşa numiţii hiperparametri) în locul calculului de integrale.

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

Definiţie. Presupunem că avem date cu funcţia de verosimilitate $f(x|\theta)$ depinzând de un parametru θ. Mai presupunem că repartiţia a priori pentru θ este una dintr-o familie de repartiţii parametrizate. Dacă repartiţia a posteriori pentru θ este în aceeaşi familie ca repartiţia a priori, spunem că a priori este o a priori conjugată pentru verosimilitate.

3.3 Repartiţia beta

Arătăm că repartiţia beta este o a priori conjugată pentru verosimilităţi binomiale, Bernoulli şi geometrice.

3.3.1 Verosimilitate binomială

Am văzut că repartiţia beta este o a priori conjugată pentru repartiţia binomială. Aceasta înseamnă că dacă funcţia de verosimilitate este binomială şi repartiţia a priori este beta, atunci repartiţia a posteriori este tot beta.
Mai concret, presupunem că funcţia de verosimilitate are o repartiţie binomială(N, θ), unde N este cunoscut şi θ este parametrul (necunoscut) de interes. Avem de asemenea că data x este un întreg între 0 şi N. Atunci, pentru o a priori beta avem următorul tabel:

hypothesis	data	prior	likelihood	posterior
θ	x	beta(a, b)	binomial(N, θ)	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 coeficienţii de normalizare ca c_1, c_2 şi c_3.

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

3.3.2 Verosimilitate Bernoulli

Repartiţia beta este o a priori conjugată pentru repartiţia Bernoulli. Acesta este de fapt un caz special al repartiţiei binomiale, deoarece Bernoulli(θ) este aceeaşi ca binomiala$(1, \theta)$. În tabelul de mai jos, arătăm actualizările corespunzând succesului $(x = 1)$ şi eşecului $(x = 0)$ pe linii separate.

hypothesis	data	prior	likelihood	posterior
θ	x	beta(a, b)	Bernoulli(θ)	beta$(a+1, b)$ or 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 şi c_3 au aceleaşi formule ca în cazul precedent (al verosimilităţii binomiale) cu $N = 1$.

3.3.3 Verosimilitate geometrică

Repartiţia geometrică(θ) descrie probabilitatea a x eşecuri 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 şi ipoteza noastră θ este că x este extrasă dintr-o repartiţie geometrică(θ). Din tabel vedem că repartiţia beta este o a priori conjugată pentru o verosimilitate geometrică:

ipoteza	data	a priori	verosimilitatea	a posteriori
θ	x	beta(a, b)	geometrică(θ)	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 şi Luigi

au găsit nişte monede neobişnuite. 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 obţine un avers în 5 aruncări.
b) Luigi decide să arunce o monedă până la primul avers. El obţine 4 reversuri înaintea primului avers.
Arătaţi că Mario şi Luigi ajung la aceeaşi a posteriori pentru θ şi calculaţi 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 şi a lui Luigi au forma unei repartiţii $\text{beta}(6,9)$. Factorul de normalizare este acelaşi în ambele cazuri deoarece este determinat cerând ca probabilitatea totală să fie 1.

3.4 Normala generează normală

Repartiţia normală este a priori conjugată cu ea însăşi. În particular, dacă funcţia de verosimilitate este normală cu dispersie cunoscută, atunci o a priori normală dă o a posteriori normală. Acum atât ipotezele şi 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 şi ştim că verosimilitatea vine dintr-o repartiţie normală cu dispersia σ^2. Dacă alegem o pdf a priori normală

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

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

$$\frac{\mu_{\text{post}}}{\sigma^2_{\text{post}}} = \frac{\mu_{\text{prior}}}{\sigma^2_{\text{prior}}} + \frac{x}{\sigma^2}, \quad \frac{1}{\sigma^2_{\text{post}}} = \frac{1}{\sigma^2_{\text{prior}}} + \frac{1}{\sigma^2}. \tag{1}$$

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

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

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

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

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

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

Răspuns.

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

Înmulţim a priori cu verosimilitatea pentru a obţine 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

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

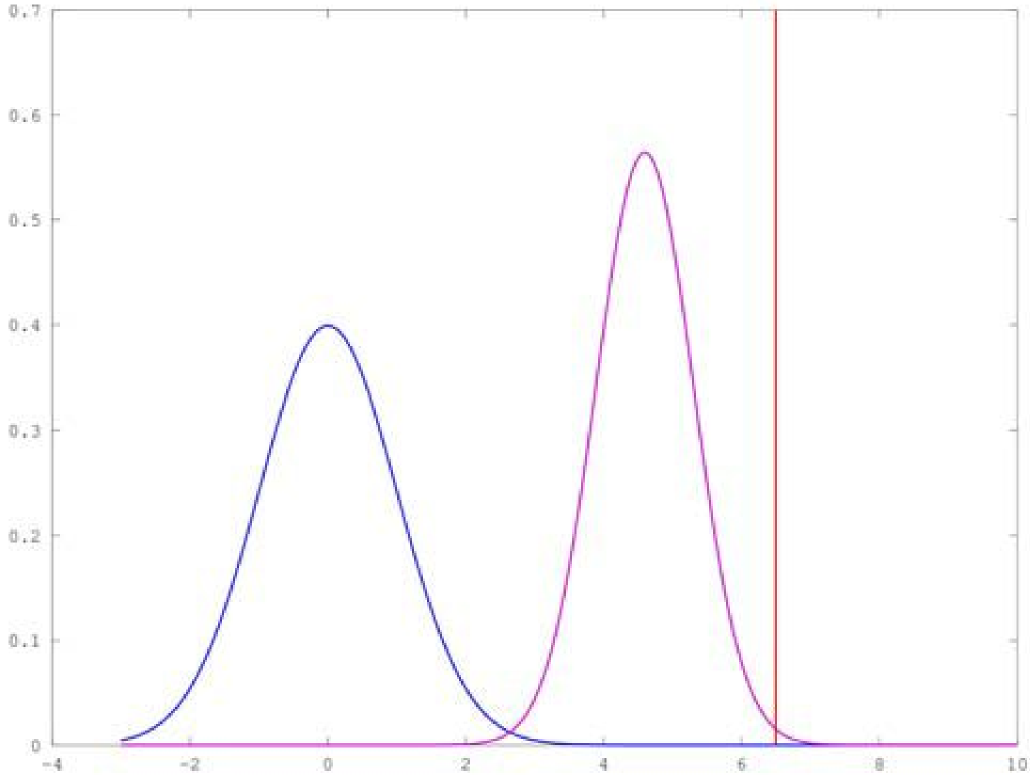
Aceasta are forma pdf pentru $N(44/13, 40/13)$, q.e.d.
Verificăm aceasta cu formulule (2).

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

De aceea

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

Exemplul 3. Presupunem că ştim datele $x \sim N(\theta, \sigma^2)$ şi avem a priori $N(0,1)$. Obţinem o valoare a datelor $x = 6.5$. Descrieţi 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 roşie.



A priori în albastru, a posteriori în mov, data în roşu.
Media a posteriori va fi o medie ponderată dintre media a priori şi dată. Vârful pdf a posteriori va fi între vârful a priori şi linia roşie. Avem

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

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. Folosiţi formulele (1) pentru a actualiza succesiv.
Răspuns. Notăm media şi dispersia a priori cu μ_0, respectiv σ_0^2. Mediile şi

dispersiile actualizate vor fi μ_i, respectiv σ_i. Avem succesiv

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

Exemplul se generalizează la n valori ale datelor $x_1, ..., x_n$:
Formule de actualizare normală-normală pentru n date

$$\frac{\mu_{\text{post}}}{\sigma_{\text{post}}^2} = \frac{\mu_{\text{prior}}}{\sigma_{\text{prior}}^2} + \frac{n\overline{x}}{\sigma^2}, \; \frac{1}{\sigma_{\text{post}}^2} = \frac{1}{\sigma_{\text{prior}}^2} + \frac{n}{\sigma^2}, \; \overline{x} = \frac{x_1 + ... + x_n}{n}. \tag{3}$$

Din nou dăm o formă mai simplu de citit, arătând că μ_{post} este o medie ponderată între μ_{prior} şi media de selecţie $\overline{x}$:

$$a = \frac{1}{\sigma_{\text{prior}}^2}, \; b = \frac{n}{\sigma^2}, \; \mu_{\text{post}} = \frac{a\mu_{\text{prior}} + b\overline{x}}{a + b}, \; \sigma_{\text{post}}^2 = \frac{1}{a + b}. \tag{4}$$

Interpretare: μ_{post} este o medie ponderată între μ_{prior} şi $\overline{x}$. Dacă numărul datelor este mare, atunci ponderea b este mare şi $\overline{x}$ va avea o puternică influenţă asupra a posteriori. Dacă σ_{prior}^2 este mică, atunci ponderea a este mare şi μ_{prior} va avea o puternică influenţă asupra a posteriori. Pentru a rezuma:
1. Multe date au o mare influenţă asupra a posteriori.
2. Siguranţa mare (dispersia mică) în a priori are o mare influenţă asupra a posteriori.