Μονοπαραμετρικά Μοντέλα

Εδώ θα θέσουμε τα θεμέλια της εκτίμησης κατά Bayes αρχίζοντας με τα μονοπαραμετρικά μοντέλα δηλαδή όταν $\vartheta : \Omega \to \Theta \subseteq \mathbb{R}$.

Εκτίμηση πιθανότητας από binomial data

Έστω δεδομένα που δίδονται με την μορφή αποτελεσμάτων δοκιμών Bernoulli. Δηλαδή τα δεδομένα μας $y = (y_1, \ldots, y_n)$ είναι τέτοια ώστε $y_i \in \{0,1\}$.

Για παράδειγμα $\{y_i = 1\}$ θα μπορούσε να σημαίνει ότι ο i ασθενής επιβιώνει πέραν του αναμενώμενου ορίου μετά την εφαρμογή κάποιας νέας θεραπείας, ενώ το ενδεχόμενο $\{y_i = 0\}$ ότι δεν επιβιώνει. Έστω ϑ η αναλογία των επιτυχιών στο πληθυσμό, όπου όλοι οι ασθενείς πάσχουν από την συγκεκριμένη ασθένεια. Το ϑ μπορεί να θεωρηθεί σαν άγνωστη πιθανότητα επιτυχίας σε ανεξάρτητες δοκιμές Bernoulli.

Επίσης δεχόμαστε ότι οι παρατηρήσεις είναι μεταξύ τους ανταλλάξιμες, στην ουσία δηλαδή δεχόμαστε την υπό συνθήκη ανεξαρτησία των παρατηρήσεων, ή ότι οι παρατηρήσεις είναι ανεξάρτητες δοθείσας της άγνωστης παραμέτρου ϑ

$$[y_i | \vartheta] \overset{iid}{\sim} Bin(\cdot|1, \vartheta) = Bernoulli(\cdot|\vartheta),\ 1 \le i \le n .$$

Ενώ ισοδύναμα για $x = \sum_{i=1}^n y_i$ έχουμε $[x|\vartheta] = \left[\sum_{i=1}^n y_i | \vartheta\right] \sim Bin(\cdot|n, \vartheta)$.

Βάση λοιπόν του μηχανισμού με τον οποίο έγινε η δειγματοληψία το μοντέλο πιθανοφάνειας (το παραμετρικό μοντέλο) είναι διωνυμικό

$$\pi_1(x|\vartheta) = Bin(x|n, \vartheta) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x},\ x \in \{0, 1, \ldots, n\} .$$

Παρατηρήστε ότι η πιθανοφάνεια χρησιμοποιώντας n Bernoulli παρατηρήσεις είναι

$$\pi_2(y_1, \ldots, y_n | \vartheta) = \prod_{i=1}^n Bin(y_i | 1, \vartheta) = \prod_{i=1}^n \vartheta^{y_i} (1-\vartheta)^{1-y_i} = \vartheta^x (1-\vartheta)^{n-x},$$

και ότι $\pi_1 \overset{\vartheta}{\propto} \pi_2$ δηλαδή οι π_1 και π_2 πιθανοφάνειες είναι $\vartheta-$ανάλογες.

<u>Παράδειγμα</u>

Ειδικοί σε ιατρικά θέματα εκφράζουν την πεποίθηση ότι η αναλογία ϑ των επιτυχιών της νέας θεραπείας στο πληθυσμό έχει τα χαρακτηριστικά

$\mathbb{E}(\vartheta)=\mu$, $Var(\vartheta)=\sigma^2$.

Η τ.μ. ϑ εμφανώς θα πρέπει να έχει στήριγμα το διάστημα $(0,1)$. Θέτουμε λοιπόν σαν μοντέλο για prior την κατανομή beta με παραμέτρους (p,q)[1], συμβατές με το γεγονός $\mathbb{E}(\vartheta)=\mu$, $Var(\vartheta)=\sigma^2$.

Μερικοί από τους λόγους για τους οποίους η beta πυκνότητα $\pi(\vartheta)=Be(\vartheta\,|\,p,q)$ είναι μια καλή επιλογή για prior είναι:

- Η κατανομή $Be(\vartheta|\,p,q)$ έχει στήριγμα το διάστημα $(0,1)$ (το ϑ είναι πιθανότητα).
- Είναι μονοκόρυφη (unimodal).
- Οι υπερπαράμετροι της (p,q) μπορούν να υπολογιστούν έτσι ώστε $E(\vartheta)=\mu$ και $Var(\vartheta)=\sigma^2$. Δηλαδή η λύση του συστήματος $p=p(\mu,\sigma^2)$, $q=q(\mu,\sigma^2)$ οδηγεί σε μοναδική λύση.

Η πυκνότητα της Beta κατανομής είναι $Be(\vartheta|\,p,q)=B(p,q)^{-1}\,\vartheta^{p-1}(1-\vartheta)^{q-1}1(0<\vartheta<1)$,

όπου $B(p,q)=\int_0^1 x^{p-1}(1-x)^{q-1}dx=\dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ το beta ολοκλήρωμα.

Για την posterior έχουμε

$$\pi(\vartheta|x)\propto\pi(\vartheta)\pi(x|\vartheta)\propto\left[\vartheta^{p-1}(1-\vartheta)^{q-1}\right]\times\left[\vartheta^{x}(1-\vartheta)^{n-x}\right]=\vartheta^{(p+x)-1}(1-\vartheta)^{(q+n-x)-1},$$

που δίνει $[\vartheta|x]\sim Be(\,\cdot\,|\,p+x,q+n-x)$

<u>Παρατηρήσεις</u>

1. **Στην περίπτωση της beta prior η posterior ανήκει στην ίδια οικογένεια κατανομών, είναι και αυτή beta**. Έτσι ο συνδυασμός της a-priori γνώσης

[1] Οι παράμετροι p,q της prior είναι οι υπερπαράμετροι του μοντέλου.

(υπό την μορφή της beta κατανομής) και των δεδομένων, χρησιμοποιώντας τον κανόνα του Bayes, έχει σαν αποτέλεσμα το update των a-priori παραμέτρων σε a-posteriori παραμέτρους. $(p,q) \to (p+x, q+n-x)$.

2. Όσο αυξάνεται ο αριθμός των παρατηρήσεων τόσο πιο κοντά είναι η, **ως προς τετραγωνική συνάρτηση απώλειας**, η σημειακή εκτίμηση κατά Bayes $\widehat{\vartheta}_{BAYES}$ (που σε αυτή τη περίπτωση είναι η posterior μέση τιμή) και η κλασική εκτίμηση $\widehat{\vartheta}_{\text{ΕΜΠ}}$ (ο εκτιμητής μέγιστης πιθανοφάνειας)

   Prior mean: $\mathbb{E}(\vartheta) = \dfrac{p}{p+q}$,

   Posterior mean: $\hat{\vartheta}_{BAYES} = \mathbb{E}(\vartheta|x) = \dfrac{p+x}{p+q+n} \approx \dfrac{x}{n} = \hat{\vartheta}_{MLE},\ n, x >> 1$.

   Ως προς γραμμική συνάρτηση απώλειας:

   Εάν $\hat{\vartheta}_{BAYES} =$ το a – posteriori $\dfrac{k_1}{k_0+k_1}$ ποσοστιαίο σημείο, τότε το $\hat{\vartheta}_{BAYES}$ θα πρέπει

   να ικανοποιεί την εξίσωση $\mathrm{B}\left(\hat{\vartheta}_{BAYES}; p+x, q+n-x\right) = \dfrac{k_1}{k_0+k_1}$, όπου

   $$\mathrm{B}(x;a,b) = \left\{\begin{matrix} \int_{\vartheta=0}^{x} Be(\vartheta \mid a,b)\, d\vartheta & x<1 \\ 1 & x \ge 1 \end{matrix}\right\},$$

   η συνάρτηση κατανομής της beta (the incomplete beta function).

   Ως **προς** $0-1$ **συνάρτηση απώλειας:**

   $\hat{\vartheta}_{BAYES} =$ a-posteriori mode (MAP) $= \dfrac{p+x-1}{p+q+n-2}$, για $p>1$ και $q>1$, εφόσον

   για $p \le 1$ και $q \le 1$ η $\pi(\vartheta)$ δεν έχει mode (ζητάμε ταυτόχρονα η $\pi(\vartheta)$ αλλά

   και η $\pi(\vartheta|x)$ να έχουν mode).

3. Όσο το μέγεθος του δείγματος αυξάνει η prior επηρεάζει όλο και λιγότερο την posterior, τελικά $Var(\vartheta|x) \to 0$ όταν $n \to \infty$.

   Prior variance:

$$Var(\vartheta) = \frac{pq}{(p+q)^2(p+q+1)},$$

Posterior variance: Εάν x και n έχουν την ίδια τάξη μεγέθους, για μεγάλο n θα έχουμε

$$Var(\vartheta|x) = \frac{(p+x)(q+n-x)}{(p+q+n)^2(p+q+n+1)} \cong \frac{1}{n} \cong 0.$$

Πάντοτε ζητάμε η posterior να έχει μεταβλητότητα μικρότερη από αυτήν της prior, δηλαδή ζητάμε η prior και το sampling distribution να είναι τέτοια ώστε να οδηγούν σε αποτελεσματική εκτίμηση ή ότι

$$Var(\vartheta|x) < Var(\vartheta)$$

Γενικότερα ισχύει ότι κατά μέση τιμή, το posterior variance είναι μικρότερο του prior variance. Πράγματι επειδή ισχύει $Var(\vartheta) = \mathbb{E}\left[Var(\vartheta|x)\right] + Var\left[\mathbb{E}(\vartheta|x)\right]$ και εφόσον $Var\left[\mathbb{E}(\vartheta|x)\right] > 0$ θα έχουμε και $\mathbb{E}\left[Var(\vartheta|x)\right] < Var(\vartheta)$.

4. Συγκρίνοντας την prior και την posterior βλέπουμε ότι η prior πυκνότητα $Be(\vartheta|p,q) \propto \vartheta^{p-1}(1-\vartheta)^{q-1}$ περιέχει «prior observations». Δηλαδή $p-1$ ψευδό-επιτυχίες (prior successes) και $q-1$ ψευδό-αποτυχίες (prior failures).

<u>Άσκηση</u>

Εάν $\vartheta \sim Be(\cdot|p,q)$ δείξτε ότι $E(\vartheta^k) = \frac{(p)_k}{(p+q)_k}$, όπου $(n)_k = n(n+1)\cdots(n+k-1)$ με $(n)_0 = 1$, (ascending or rising factorial). Επίσης δείξτε ότι $Var(\vartheta) = pq\left\{(p+q)^2(p+q+1)\right\}^{-1}$.

<u>Λύση</u>

$$\mathbb{E}(\vartheta^k) = \int_0^1 x^k Be(x|p,q)\,dx = \frac{1}{B(p,q)}\int_0^1 x^{p+k-1}(1-x)^{q-1}\,dx = \frac{B(p+k,q)}{B(p,q)}$$

$$= \left\{\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\right\}\left\{\frac{\Gamma(p+k)\Gamma(q)}{\Gamma(p+q+k)}\right\} = \left\{\frac{\Gamma(p+q)}{\Gamma(p)}\right\}\left\{\frac{\Gamma(p+k)}{\Gamma(p+q+k)}\right\}$$

$$=\left\{\frac{(p+k-1)\cdots p\Gamma(p)}{\Gamma(p)}\right\}\left\{\frac{\Gamma(p+q)}{(p+q+k-1)\cdots(p+q)\Gamma(p+q)}\right\}$$

$$=\frac{p(p+1)\cdots(p+k-1)}{(p+q)(p+q+1)\cdots(p+q+k-1)}=\frac{(p)_k}{(p+q)_k}.$$

$$Var(\vartheta)=\mathbb{E}\left(\vartheta^2\right)-\mathbb{E}(\vartheta)^2=\frac{p(p+1)}{(p+q)(p+q+1)}-\frac{p^2}{(p+q)^2}=\frac{pq}{(p+q)^2(p+q+1)}.$$

<u>Άσκηση</u>

Δείξτε ότι η κατανομή $\vartheta \sim Be(p,q)$ έχει mode $\vartheta_{Mode}=\frac{p-1}{p+q-2}$ μόνο όταν $p>1$ και $q>1$.

$$\frac{\partial}{\partial\vartheta}\log Be(\vartheta\,|\,p,q)=0 \Rightarrow \vartheta=\frac{p-1}{p+q-2}$$

$$\frac{\partial^2}{\partial\vartheta^2}\log Be(\vartheta\,|\,p,q)<0 \Rightarrow p>1, q>1.$$

<u>Άσκηση</u>

Δείξτε ότι $Cov(\vartheta,\varphi)=\mathbb{E}\left[Cov(\vartheta,\varphi|x)\right]+Cov\left[\mathbb{E}(\vartheta|x),\mathbb{E}(\varphi|x)\right]$, καθώς και το πιο ειδικό αποτέλεσμα $Var(\vartheta)=\mathbb{E}\left[Var(\vartheta|x)\right]+Var\left[\mathbb{E}(\vartheta|x)\right]$.

$$\mathbb{E}\left[Cov(\vartheta,\varphi|x)\right]=\mathbb{E}\left[\mathbb{E}(\vartheta\varphi|x)-\mathbb{E}(\vartheta|x)E(\varphi|x)\right]$$

$$=\mathbb{E}(\vartheta\varphi)-\mathbb{E}\left[\mathbb{E}(\vartheta|x)\mathbb{E}(\varphi|x)\right]$$

$$Cov\left[\mathbb{E}(\vartheta|x),\mathbb{E}(\varphi|x)\right]=\mathbb{E}\left[\mathbb{E}(\vartheta|x)\mathbb{E}(\varphi|x)\right]-\mathbb{E}\left[\mathbb{E}(\vartheta|x)\right]\mathbb{E}\left[\mathbb{E}(\varphi|x)\right]$$

$$=\mathbb{E}\left[\mathbb{E}(\vartheta|x)\mathbb{E}(\varphi|x)\right]-\mathbb{E}[\vartheta]\mathbb{E}[\varphi]$$

Στη συνέχεια προσθέτουμε κατά μέλη και έχουμε:

$$Cov(\vartheta,\varphi)=\mathbb{E}\left[Cov(\vartheta,\varphi|x)\right]+Cov\left[\mathbb{E}(\vartheta|x),\mathbb{E}(\varphi|x)\right].$$

Επειδή $Var(\vartheta)=Cov(\vartheta,\vartheta)$, παίρνουμε $Var(\vartheta)=\mathbb{E}\left[Var(\vartheta|x)\right]+Var\left[\mathbb{E}(\vartheta|x)\right]$.

Εναλλακτικά η εξίσωση για το $Var(\vartheta)$ μπορεί να αποδειχτεί με παρόμοιο τρόπο:

$$\mathbb{E}\left[Var(\vartheta|x)\right]=\mathbb{E}\left[\mathbb{E}(\vartheta^2|x)-\mathbb{E}(\vartheta|x)^2\right]=\mathbb{E}(\vartheta^2)-\mathbb{E}\left[\mathbb{E}(\vartheta|x)^2\right]$$

$$Var\left[\mathbb{E}(\vartheta|x)\right]=\mathbb{E}\left[\mathbb{E}(\vartheta|x)^2\right]-\mathbb{E}\left[\mathbb{E}(\vartheta|x)\right]^2=\mathbb{E}\left[\mathbb{E}(\vartheta|x)^2\right]-\mathbb{E}(\vartheta)^2.$$

Αριθμητικό παράδειγμα

Έστω ότι $n=70$ με διωνυμική παρατήρηση $x=34$. Εάν οι πεποιθήσεις των ειδικών είναι ότι $\mu=0.4$ και $\sigma^2=0.02$, οι παράμετροι p,q της beta prior $Be(\vartheta|p,q)$ θα ικανοποιούν το μη γραμμικό 2×2 σύστημα:

$$\left\{\begin{array}{l}\dfrac{p}{p+q}=\mu\\ \dfrac{pq}{(p+q)^2(p+q+1)}=\sigma^2\end{array}\right\}.$$

Από την πρώτη εξίσωση έχουμε $p+q=\dfrac{p}{\mu},\ p+q+1=\dfrac{p+\mu}{\mu}$ και αντικαθιστώντας στη δεύτερη εξίσωση

$$\frac{pq}{\left(\frac{p}{\mu}\right)^2\left(\frac{p+\mu}{\mu}\right)}=\sigma^2 \Leftrightarrow q=\frac{\sigma^2}{\mu^3}p(p+\mu)$$

$$p+q=\frac{p}{\mu} \Leftrightarrow p+\frac{\sigma^2}{\mu^3}p(p+\mu)=\frac{p}{\mu} \Leftrightarrow p=\frac{(1-\mu)\mu^2}{\sigma^2}-\mu$$

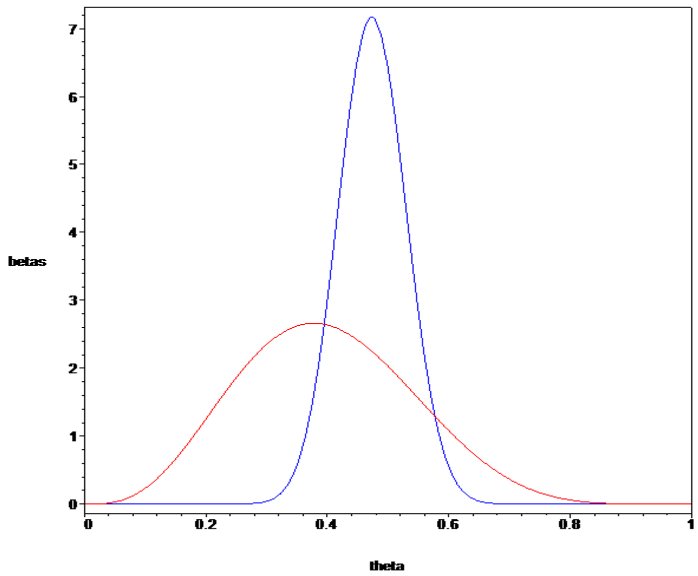
$$p+q=\frac{p}{\mu} \Leftrightarrow q=\left(\frac{1}{\mu}-1\right)p=\left(\frac{1}{\mu}-1\right)\left(\frac{(1-\mu)\mu^2}{\sigma^2}-\mu\right)=\frac{(1-\mu)^2\mu}{\sigma^2}-(1-\mu),$$

Τελικά $\left\{\begin{array}{l}p=\dfrac{(1-\mu)\mu^2}{\sigma^2}-\mu\\ q=\dfrac{\mu(1-\mu)^2}{\sigma^2}-(1-\mu)\end{array}\right\}$.

Για $(\mu,\sigma^2)=(0.4,0.02)$ παίρνουμε $(p,q)=(4.4,6.6)$.

Έτσι η προηγούμενη ανάλυση μας δίνει

$\vartheta \sim Be(4.4,6.6) \Rightarrow [\vartheta | x=34] \sim Be(38.4,42.6)$



$\mathbb{E}(\vartheta)=0.400$,

$\mathbb{E}[\vartheta | x=34]=\widehat{\vartheta}_{Bayes}=0.474$,

$\widehat{\vartheta}_{MLE}=0.486$ με απόκλιση 2.47% [2]

$Var[\vartheta]=0.020$

$Var[\vartheta | x=34]=0.003$.

Έχουμε κάνει λοιπόν μια αποτελεσματική εκτίμηση της παραμέτρου ϑ εφόσον $Var[\vartheta | x=34] < Var(\vartheta)$.

[2] To percentage absolute relative error του $\widehat{\vartheta}_{MLE}$ ως προς το $\widehat{\vartheta}_{Bayes}$ είναι $100\left|\dfrac{\widehat{\vartheta}_{MLE}-\widehat{\vartheta}_{Bayes}}{\widehat{\vartheta}_{MLE}}\right|$

Έστω τώρα ότι έρχεται και δεύτερη διωνυμική παρατήρηση $x_2 = 37$ από δεύτερο δείγμα Bernoulli μεγέθους και αυτό $n_2 = 70$.

Κάνοντας διαδοχική ανάλυση με

prior $[\vartheta|x_1] \sim Be(\cdot|p+x_1, q+n_1-x_1)$, θα έχουμε νέα πιθανοφάνεια $[x_2|\vartheta] \sim Bin(\cdot|n_2,\vartheta)$, και posterior $[\vartheta|x_1,x_2] \sim Be(\cdot|p+(x_1+x_2), q+(n_1+n_2)-(x_1+x_2))$.

πράγματι

$$\pi(\vartheta|x_1,x_2) \propto \pi(\vartheta,x_1,x_2) \propto \pi(x_1)\pi(\vartheta|x_1)\pi(x_2|\vartheta,x_1)$$

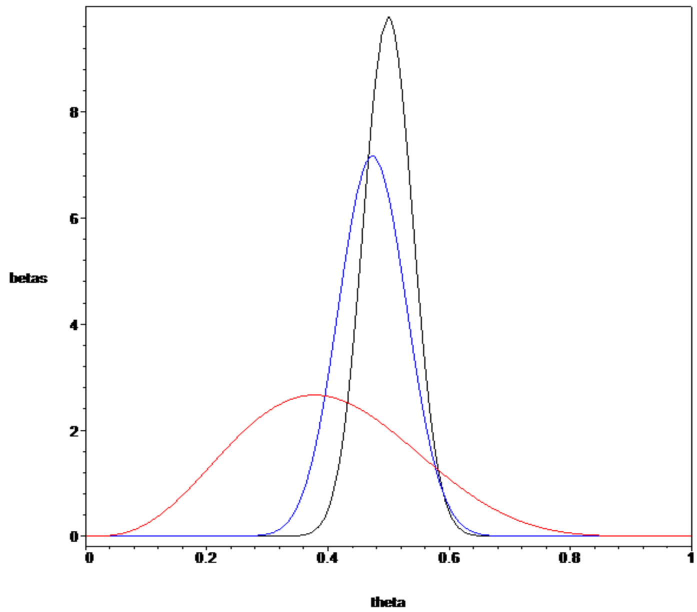
$$\propto \pi(\vartheta|x_1)\pi(x_2|\vartheta) \propto \left\{\vartheta^{(p+x_1)-1}(1-\vartheta)^{(q+n_1-x_1)-1}\right\} \times \left\{\vartheta^{x_2}(1-\vartheta)^{n_2-x_2}\right\}$$

$$= \vartheta^{(p+x_1+x_2)-1}(1-\vartheta)^{(q+n_1+n_2-x_1-x_2)-1} \propto Be(\cdot|p+(x_1+x_2), q+(n_1-x_1)+(n_2-x_2))$$

κάνοντας αριθμητική αντικατάσταση έχουμε

$$\vartheta \sim Be(\cdot|4.4,6.6) \Rightarrow [\vartheta|x=34] \sim Be(38.4,42.6)$$

$$\Rightarrow [\vartheta|x_1=34, x_2=37] \sim Be(75.4,75.6)$$



$\mathbb{E}[\vartheta|x_1=34] = 0.474$

$\mathbb{E}[\vartheta|x_1=34, x_2=37] = \hat{\vartheta}_{Bayes} = 0.499$, $\hat{\vartheta}_{MLE} = 0.507$ με απόκλιση 1.58%

$Var\left[\vartheta | x_1 = 34\right] = 0.003$,

$Var\left[\vartheta | x_1 = 34, x_2 = 37\right] = 0.002$.

Έχουμε και πάλι κάνει αποτελεσματική εκτίμηση της παραμέτρου ϑ εφόσον

$Var\left[\vartheta | x_1 = 34,\ x_2 = 37\right] < Var\left[\vartheta | x_1 = 34\right] < Var\left[\vartheta\right]$.

Η prior predictive για το beta – binomial μοντέλο για μια διωνυμική παρατήρηση x είναι η μίξη της πιθανοφάνειας $\pi\left(x|\vartheta\right)$ ως προς το prior μέτρο $\Pi\left(d\vartheta\right) = Be\left(\vartheta | p,q\right) d\vartheta$.

$$\pi\left(x\right) = \int_{\Theta} \pi\left(x|\vartheta\right) \pi\left(\vartheta\right) d\vartheta = \int_{\vartheta=0}^{1} Bin\left(x|n,\vartheta\right) Be\left(\vartheta | p,q\right) d\vartheta$$

$$= \binom{n}{x} \frac{\Gamma\left(p+q\right)}{\Gamma\left(p\right)\Gamma\left(q\right)} \int_0^1 \vartheta^{p+x-1} \left(1-\vartheta\right)^{q+n-x-1} d\vartheta = \binom{n}{x} \frac{\Gamma\left(p+q\right)}{\Gamma\left(p\right)\Gamma\left(q\right)} \left\{ \frac{\Gamma\left(p+x\right)\Gamma\left(q+n-x\right)}{\Gamma\left(p+q+n\right)} \right\}$$

$$= \binom{n}{x} \frac{B\left(p+x, q+n-x\right)}{B\left(p,q\right)},\ x \in \left\{0,1,\ldots,n\right\}$$

Η $\pi\left(x\right)$ είναι μία **σύνθετη κατανομή** (compound probability distribution), προκύπτει δηλαδή από μίξη, κατά την έννοια ότι η πιθανότητα επιτυχίας ϑ, στην διωνυμική κατανομή $Bin\left(n, \vartheta\right)$, προέρχεται από την $Be\left(p,q\right)$**. Σε αυτή την περίπτωση λέμε ότι η prior predictive ακολουθεί την beta – binomial κατανομή**:

$x \sim BB\left(n,p,q\right)$,

$$BB\left(x | n,p,q\right) = \left\{ \begin{matrix} \binom{n}{x} \frac{\Gamma\left(p+q\right)}{\Gamma\left(p\right)\Gamma\left(q\right)} \left\{ \frac{\Gamma\left(p+x\right)\Gamma\left(q+n-x\right)}{\Gamma\left(p+q+n\right)} \right\} & x \in \left\{0,1,\ldots,n\right\} \\ 0 & elsewhere \end{matrix} \right\}.$$

Εναλλακτικά θα μπορούσαμε, εφόσον γνωρίζουμε ακριβώς τις ποσότητες $\pi\left(\vartheta|x\right)$, $\pi\left(\vartheta\right)$ και $\pi\left(x|\vartheta\right)$, να υπολογίσουμε την $\pi\left(x\right)$ από τον κανόνα του Bayes. Έτσι για το $\pi\left(x\right)$ έχουμε:

$$\pi(x)=\frac{\pi(\vartheta)\pi(x|\vartheta)}{\pi(\vartheta|x)}=\frac{Be(\vartheta|p,q)Bin(x|n,\vartheta)}{Be(\vartheta|p+x,q+n-x)}$$

$$=\frac{\dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\vartheta^{p-1}(1-\vartheta)^{q-1}\binom{n}{x}\vartheta^{x}(1-\vartheta)^{n-x}}{\dfrac{\Gamma(p+q+n)}{\Gamma(p+x)\Gamma(q+n-x)}\vartheta^{p+x-1}(1-\vartheta)^{q+n-x-1}}=\binom{n}{x}\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\left\{\frac{\Gamma(p+x)\Gamma(q+n-x)}{\Gamma(p+q+n)}\right\}.$$

<u>Άσκηση</u>

1. Εάν $x\sim BB(n,p,q)$, να βρεθεί η ροπογεννήτρια συνάρτηση $M_x(t)$ και στη συνέχεια η μέση τιμή και η διασπορά της x.
2. Δείξτε ότι η $BB(1,p,q)=Bin\left(1,\dfrac{p}{p+q}\right)$.

1. $M_x(t)=\sum_{x=0}^{n}e^{tx}BB(x|n,p,q)=\sum_{x=0}^{n}e^{tx}\int_0^1\binom{n}{x}\vartheta^{x}(1-\vartheta)^{n-x}\dfrac{\vartheta^{p-1}(1-\vartheta)^{q-1}}{B(p,q)}d\vartheta$

$$=\int_0^1\sum_{x=0}^{n}\binom{n}{x}(\vartheta e^{t})^{x}(1-\vartheta)^{n-x}\frac{\vartheta^{p-1}(1-\vartheta)^{q-1}}{B(p,q)}d\vartheta=\int_0^1(\vartheta e^{t}+1-\vartheta)^{n}\frac{\vartheta^{p-1}(1-\vartheta)^{q-1}}{B(p,q)}d\vartheta$$

$$M_x^{(1)}(0)=\int_0^1(n\vartheta)\frac{\vartheta^{p-1}(1-\vartheta)^{q-1}}{B(p,q)}d\vartheta=\frac{n}{B(p,q)}\int_0^1\vartheta^{(p+1)-1}(1-\vartheta)^{q-1}d\vartheta=n\frac{B(p+1,q)}{B(p,q)}=n\frac{p}{p+q},$$

$$M_x^{(2)}(0)=\int_0^1\left[n\vartheta+n(n-1)\vartheta^{2}\right]\frac{\vartheta^{p-1}(1-\vartheta)^{q-1}}{B(p,q)}d\vartheta=n\frac{B(p+1,q)}{B(p,q)}+n(n-1)\frac{B(p+2,q)}{B(p,q)}$$

$$=\frac{np}{p+q}+\frac{n(n-1)p(p+1)}{(p+q)(p+q+1)}.$$

Οι προηγούμενες σχέσεις δίνουν

$$\mathbb{E}(x)=\frac{np}{p+q},\quad Var(x)=\frac{npq(p+q+n)}{(p+q)^{2}(p+q+1)}.$$

2. $BB(x|1,p,q)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\left\{\frac{\Gamma(p+x)\Gamma(q+1-x)}{\Gamma(p+q+1)}\right\},\ x\in\{0,1\}$

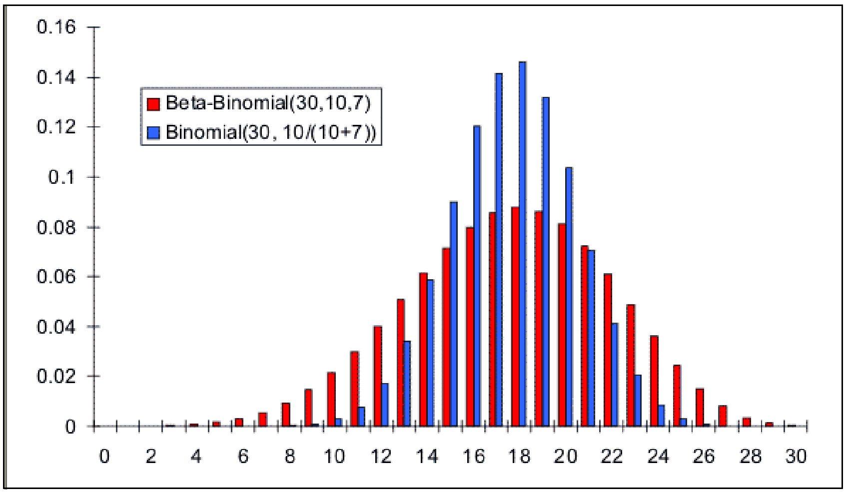
$$=\left\{\begin{array}{l}\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\left\{\frac{\Gamma(p)\Gamma(q+1)}{\Gamma(p+q+1)}\right\}=\frac{q}{p+q},\ x=0\\ \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\left\{\frac{\Gamma(p+1)\Gamma(q)}{\Gamma(p+q+1)}\right\}=\frac{p}{p+q},\ x=1\end{array}\right\}.$$

<u>Άσκηση</u>

Η πιο καλά προσαρμοσμένη διωνυμική τ.μ. y, στην $x\sim BB(n,p,q)$, είναι η $y\sim Bin(n,p^*)$ με $p^*=p/(p+q)$. Δείξτε ότι για $\forall n>1,\ Var(y)<Var(x)$.

Έχουμε ότι:

$$Var(y)=np^*(1-p^*)=\frac{npq}{(p+q)^2}<\frac{npq(p+q+n)}{(p+q)^2(p+q+1)}=Var(x),\ \forall n>1$$



<u>Παρατήρηση</u> η σχέση για την ροπογεννήτρια συνάρτηση προκύπτει κατ' ευθείαν από την σχέση $M_x(t)=\mathbb{E}\left[M_{x|\vartheta}(t)\right]$

$$M_x(t)=\mathbb{E}\left[e^{tx}\right]=\mathbb{E}\left[\mathbb{E}\left[e^{tx}|\vartheta\right]\right]=\mathbb{E}\left[M_{x|\vartheta}(t)\right]=\int_0^1 M_{x|\vartheta}(t)\pi(\vartheta)d\vartheta$$

$$= \int_0^1 \left(\vartheta e^t + 1 - \vartheta \right)^n \left\{ \frac{\vartheta^{p-1}(1-\vartheta)^{q-1}}{B(p,q)} d\vartheta \right\}$$

<u>Πιο γενικά</u>

$$\mathbb{E}\left[M_{X|\Theta}(t) \right] = \mathbb{E}\left[\mathbb{E}\left(e^{tX} | \Theta \right) \right] = \mathbb{E}\left[\int_X e^{tx} \pi(x|\Theta) dx \right] = \int_X e^{tx} \mathbb{E}\left[\pi(x|\Theta) \right] dx$$

$$= \int_X e^{tx} \left\{ \int_\Theta \pi(x|\vartheta) \pi(\vartheta) d\vartheta \right\} dx = \int_X e^{tx} \left\{ \int_\Theta \pi(x,\vartheta) d\vartheta \right\} dx = \int_X e^{tx} \pi(x) dx$$

$$= \mathbb{E}\left(e^{tx} \right) = M_X(t).$$

<u>Άσκηση</u>

Για το Binomial – Beta μοντέλο να βρεθεί η από κοινού κατανομή των ϑ και x.

<u>Παρατήρηση</u>

The Beta-Binomial distribution always has more spread (variance) than its best fitting Binomial distribution, because the Beta distribution adds extra randomness. Thus, when a Binomial distribution does not match observations, because the observations exhibit too much spread, a Beta-Binomial distribution is often used instead.

The number of life insurance policy holders who will die in any one year, where some external variable (e.g. highly contagious disease, extreme weather) moderates the probability of death of all individual to some degree.

Η posterior predictive για το beta – binomial μοντέλο για μια νέα διωνυμική παρατήρηση y είναι η μίξη της πιθανοφάνειας $\pi(y|\vartheta)$ ως προς το posterior μέτρο $\Pi(d\vartheta|x) = Be(\vartheta|p+x, q+n-x)d\vartheta$.

Ειδικότερα εάν η y είναι νέα δοκιμή Bernoulli έχουμε:

$$\pi(y|x) = \int_\Theta \pi(y,\vartheta|x) d\vartheta = \int_\Theta \pi(y|\vartheta)\pi(\vartheta|x) d\vartheta$$

$$= \int_0^1 Bin(y|1,\vartheta) Be(\vartheta|p+x, q-x+n) d\vartheta$$

$$= BB(y|1, p+x, q+n-x) = \binom{1}{y} \frac{B((p+x)+y, (q+n-x)+1-y)}{B(p+x, q+n-x)}$$

$$= \left\{ \frac{\Gamma(p+x+y)\Gamma(q+n-x-y+1)}{\Gamma(p+q+n+1)} \right\} \left\{ \frac{\Gamma(p+q+n)}{\Gamma(p+x)\Gamma(q+n-x)} \right\}, \ y \in \{0,1\}.$$

Η πιθανότητα του $\{y=0\}$ είναι

$$\pi(y=0|x) = \left\{ \frac{\Gamma(p+x)\Gamma(q+n-x+1)}{\Gamma(p+q+n+1)} \right\} \left\{ \frac{\Gamma(p+q+n)}{\Gamma(p+x)\Gamma(q+n-x)} \right\} = \frac{q+n-x}{p+q+n}.$$

Έτσι για την posterior predictive $[y|x]$ έχουμε:

$$[y|x] \sim BB(1, p+x, q+n-x) \sim \begin{pmatrix} 0 & 1 \\ \dfrac{q+n-x}{p+q+n} & \dfrac{p+x}{p+q+n} \end{pmatrix}.$$

<u>Εναλλακτικά</u>, επειδή γνωρίζουμε ότι $BB(1, p, q) = Bin\left(1, \dfrac{p}{p+q}\right)$, θα έχουμε

$$[y|x] \sim BB(1, p+x, q+n-x) = Bin\left(1, \frac{p+x}{p+q+n}\right).$$

Δηλαδή η posterior predictiveγια μια νέα Bernoulli παρατήρηση είναι και αυτή Bernoulli.

Το posterior predictive για m νέες παρατηρήσεις Bernoulli $(x_1', \ldots, x_m')$ και $[y|\vartheta] = \left[\sum_{i=1}^{m} x_i' |\vartheta\right] \sim Bin(m, \vartheta)$ είναι:

$$\pi(y|x) = \int_0^1 Bin(y|m, \vartheta) Be(\vartheta| p+x, q-x+n) d\vartheta$$

$$= Bb(y|m, p+x, q+n-x) = \binom{m}{x'} \frac{B((p+x)+y, (q+n-x)+m-y)}{B(p+x, q+n-x)}, \ \ y \in \{0,1,\ldots,m\}.$$

A none informative prior (a flat prior) for Binomial data

Το Beta – Binomial μοντέλο κάτω από απουσία αρχικών πεποιθήσεων: Θεωρούμε και πάλι το προηγούμενο μοντέλο με binomial data μόνο που αυτή την φορά δεν διαθέτουμε την αρχική γνώση των ειδικών για την μέση τιμή και την διασπορά του ϑ. Για την Bayesian εκτίμηση όμως χρειαζόμαστε prior με στήριγμα το $(0,1)$, που όμως «δεν θα μεταφέρει» καμία a-priori πληροφορία στη posterior.

Χρησιμοποιούμε λοιπόν για αυτόν τον λόγο σαν prior την ομοιόμορφη κατανομή στο διάστημα $(0,1)$ εφόσον $P_\vartheta([a,b]) = P_\vartheta([c,d])$ για κάθε ισομήκες ζεύγος υποσυνόλων του $(0,1)$

$\vartheta \sim \mathcal{U}(0,1) \Leftrightarrow \pi(\vartheta) = 1(0<\vartheta<1) \Leftrightarrow \pi(\vartheta) \propto 1,\ 0<\vartheta<1$.

Δεν χρειαζόμαστε νέους υπολογισμούς γιατί στην ειδική περίπτωση που $p=q=1$ έχουμε $Be(\vartheta|1,1) = \mathcal{U}(\vartheta|0,1)$. Έτσι για ομοιόμορφη prior $[\vartheta|x] \sim Be(\cdot|x+1, n-x+1)$

- prior predictive $x \sim BB(n,1,1)$

  $$\pi(x) = BB(x|n,1,1) = \binom{n}{x}\frac{B(x+1,n-x+1)}{B(1,1)} = \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)} = \frac{1}{n+1}$$

  για όλα τα $x \in \{0,1,\dots,n\}$. Και όπως αναμενόταν, όλες οι δυνατές τιμές του x, a-priori είναι ισοπίθανες με $BB(n,1,1) = D\mathcal{U}\{0,1,\dots,n\}$.

  Εναλλακτικά

  $$\pi(x) = \int_0^1 Bin(x|n,\vartheta)\mathcal{U}(\vartheta|0,1)\,d\vartheta = \binom{n}{x}\int_0^1 \vartheta^x(1-\vartheta)^{n-x}\,d\vartheta$$

  $$= \binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)} = \frac{1}{n+1}.$$

- Η Posterior predictive για μια Bernoulli παρατήρηση y είναι

  $[y|x] \sim BB(1, x+1, n-x+1)$. Επειδή

  $$\pi(y=0|x) = \frac{n-x+1}{n+2},\ \pi(y=1|x) = \frac{x+1}{n+2}.$$

Δειγματοληψία από μίξη κατανομών

Έχουμε το παρακάτω σχήμα δειγματοληψίας:

$$\begin{aligned} \vartheta &\sim \pi(\,\cdot\,) = Be(\,\cdot \mid p,q) \\ y \mid \vartheta &\sim \pi(\,\cdot \mid \vartheta) \end{aligned} \quad \overset{marginally}{\Longrightarrow} \quad y \sim \pi^*(\,\cdot\,) = \int_\Theta \pi(\,\cdot \mid \vartheta)\pi(\vartheta)\,d\vartheta .$$

Για παράδειγμα

$$\begin{aligned} \vartheta &\sim Be(\,\cdot \mid p,q) \\ y \mid \vartheta &\sim Bin(\,\cdot \mid n,\vartheta) \end{aligned} \quad \overset{marginally}{\Longrightarrow} \quad \boxed{\begin{aligned} y \sim \pi^*(\,\cdot\,) &= \int_\Theta Bin(\,\cdot \mid n,\vartheta)Be(\vartheta \mid p,q)\,d\vartheta \\ &= BB(\,\cdot \mid n,p,q) \end{aligned}}$$

Η παραπάνω σχέση μας λέει ότι για να προσομοιώσουμε iid δείγμα $\left\{ y_i \overset{iid}{\sim} \pi^*(\,\cdot\,), \leq i \leq n \right\}$ θα πρέπει πρώτα να προσομοιώσουμε διάνυσμα $(\vartheta_1,\ldots,\vartheta_n)$ με $\vartheta_i \overset{iid}{\sim} \pi(\,\cdot\,)$, $1 \leq i \leq n$ και μετά διάνυσμα $(y_1,\ldots,y_n)$ με $y_i \mid \vartheta_i \overset{ind}{\sim} \pi(\,\cdot \mid \vartheta_i)$, $1 \leq i \leq n$.

Πράγματι έστω ότι οι πραγματοποιήσεις της τυχαίας μεταβλητής $\tilde{Y}$ είναι τα y στα αριστερά του σχήματος δειγματοληψίας και οι πραγματοποιήσεις της τυχαίας μεταβλητής Y^* είναι τα y στα δεξιά του σχήματος δειγματοληψίας θα δείξουμε ότι οι τ.μ. $\tilde{Y}$ και Y^* έχουν την ίδια κατανομή, συμβολικά $\tilde{Y} \overset{d}{=} Y^*$.

Η συνάρτηση κατανομής της τ.μ. $\tilde{Y}$ είναι $F_{\tilde{Y}}(y) = P\left\{\tilde{Y} \leq y\right\}$

$$F_{\tilde{Y}}(y) = P\left\{\tilde{Y} \leq y\right\} = \int_\Theta P\left\{\tilde{Y} \leq y \mid \vartheta\right\}\pi(\vartheta)\,d\vartheta$$

$$= \int_\Theta \left(\int_{u=-\infty}^{y} \pi(u \mid \vartheta)\,du \right) \pi(\vartheta)\,d\vartheta = \int_{-\infty}^{y} \left(\int_\Theta \pi(\vartheta)\pi(u \mid \vartheta)\,d\vartheta \right) du$$

Όμως $\int_\Theta \pi(\vartheta)\pi(u \mid \vartheta)\,d\vartheta$ είναι η μίξη $\pi^*(u)$ και έτσι

$$F_{\tilde{Y}}(y) = \int_{-\infty}^{y} \pi^*(u)\,du = P\left\{Y^* \leq y\right\} = F_{Y^*}(y)$$

δηλαδή έχουμε $F_{\tilde{Y}}(y) = F_{Y^*}(y)$ ή ότι $\tilde{Y} \overset{d}{=} Y^*$.

<u>Άσκηση</u>

Έστω ότι $Y \sim \mathcal{U}(0,1)$ και ότι η τ.μ. Z έχει συνάρτηση κατανομής F_Z που αντιστρέφεται. Τότε η τ.μ. $X = F_Z^{-1}(Y)$ έχει την ίδια κατανομή με την τ.μ. Z, συμβολικά $F_Z^{-1}(Y) \stackrel{d}{=} Z$.

Δείξτε ότι ισχύει και το αντίστροφο του προηγούμενου, δηλαδή ότι εάν η τ.μ. Z έχει συνάρτηση κατανομής F_Z που αντιστρέφεται. Τότε η τ.μ. $Y = F_Z(Z)$ είναι ομοιόμορφη στο $(0,1)$.

1. Για να δείξουμε ότι $X \stackrel{d}{=} Z$ αρκεί να δείξουμε ότι $F_X(x) = F_Z(x)$.

$F_X(x) = P\{X \le x\} = P\left\{F_Z^{-1}(Y) \le x\right\}$,

και επειδή F_Z $\uparrow$ παίρνουμε $F_X(x) = P\left\{Y \le F_Z(x)\right\} = \int_{-\infty}^{F_Z(x)} f_Y(y)\,dy$.

$Y \sim \mathcal{U}(0,1) \Leftrightarrow f_Y(y) = 1(0 < y < 1)$, και $0 \le F_Z(x) \le 1$ δίνει

$$F_X(x) = \int_{-\infty}^{F_Z(x)} 1(0 < y < 1)\,dy = \int_{0}^{F_Z(x)} dy = F_Z(x).$$

2. $f_Y(y) = f_Z\left(F_Z^{-1}(y)\right)\left|\left(F_Z^{-1}\right)'(y)\right| = f_Z\left(F_Z^{-1}(y)\right)\left|\left(F_Z\right)'\left(F_Z^{-1}(y)\right)\right|^{-1}$

$= f_Z\left(F_Z^{-1}(y)\right)\left|f_Z\left(F_Z^{-1}(y)\right)\right|^{-1} = 1$

εφόσον $f_Z > 0$ και $F_Z : \overline{R} \to [0,1]$ είναι ένα – προς – ένα παίρνουμε ότι

$f_Y(y) = 1(0 < y < 1)$ ή ότι $Y \sim \mathcal{U}(0,1)$.

Διαστήματα υψηλής a – posteriori πυκνότητας. Θα μπορούσε να πει κανείς, ότι τα HPDI (Highest Posterior Density Intervals ή Credible intervals)[3] είναι τα αντίστοιχα

[3] Highest Posterior Density Regions ή Credible Regions όταν $\Theta \subseteq \mathbb{R}^d$.

διαστήματα εμπιστοσύνης της κλασσικής Στατιστικής. Τα διαστήματα εμπιστοσύνης $I_a(X)=\left[S_-(X),\, S_+(X)\right]$ είναι τυχαία και περιέχουν την **σταθερή** άγνωστη παράμετρο με πιθανότητα $P\left\{S_-(X)\le \vartheta \le S_+(X)\,|\,\vartheta\right\}=1-a$. Από την στιγμή όμως που θα δούμε την πραγματοποίηση του ενδεχομένου $\{X=x\}$, η προηγούμενη πιθανότητα καταρρέει σε 0 ή 1, δηλαδή έχουμε ότι

$$P\left\{S_-(x)\le \vartheta \le S_+(x)\,|\,\vartheta\right\}=1\left(S_-(x)\le \vartheta \le S_+(x)\right),$$

και δεν έχουμε καμία περαιτέρω πληροφορία από το διάστημα εμπιστοσύνης.

Ένα κατά Bayes «διάστημα εμπιστοσύνης» είναι αριθμητικό δηλαδή:

$$P\left\{S_-(x)\le \vartheta \le S_+(x)\,|\,x\right\}=\int_{S_-(x)}^{S_+(x)}\pi(\vartheta\,|\,x)\,d\vartheta=1-a.$$

Το σύνολο $I_a(x)=\left[S_-(x),\, S_+(x)\right]\subset\Theta$, με την ιδιότητα $P\left\{\vartheta\in I_a(x)\,|\,x\right\}=1-a$ δεν είναι μοναδικό, για αυτό τον λόγο **ορίζουμε σαν HPDI** το σύνολο $I_a(x)$ που επιπροσθέτως ικανοποιεί και την συνθήκη

$$\vartheta\in I_a(x) \text{ και } \vartheta'\notin I_a(x) \;\Rightarrow\; \pi(\vartheta\,|\,x)>\pi(\vartheta'\,|\,x).$$

Στην μονοδιάστατη περίπτωση θέτουμε:

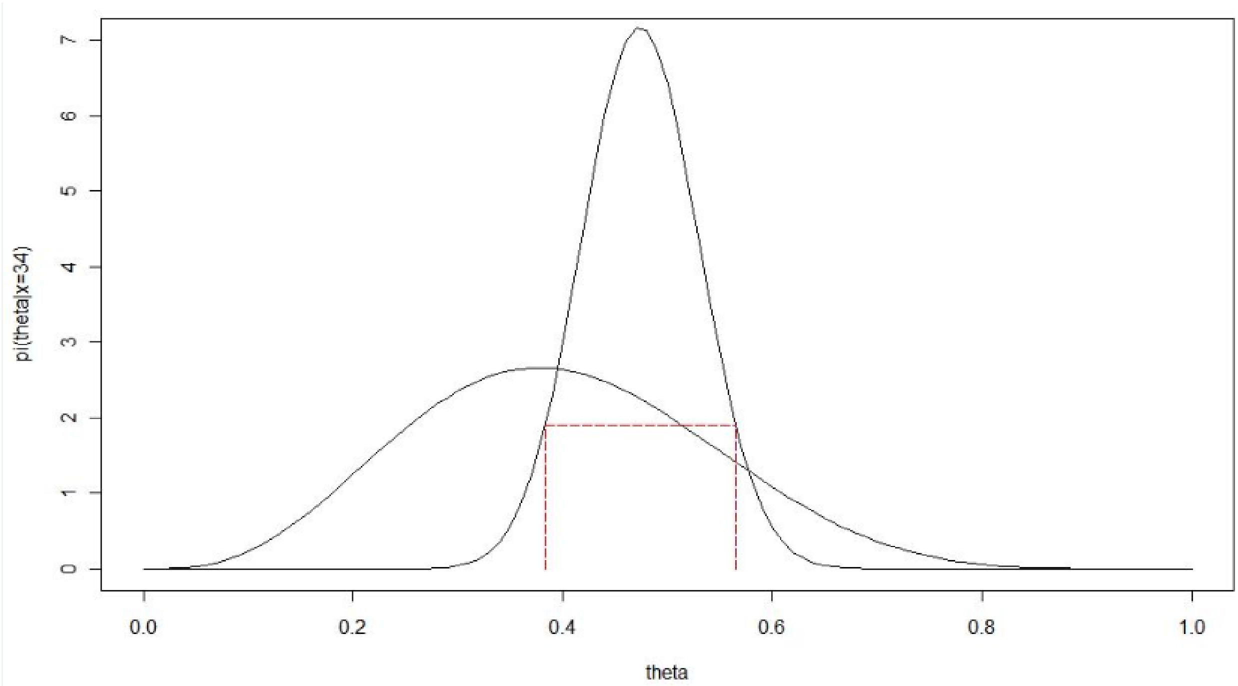
$$P\left\{\vartheta\in I_a(x)\,|\,x\right\}=1-a,\;\; I_a(x)=\left\{\vartheta\in\Theta:\pi(\vartheta\,|\,x)\ge\gamma_a>0\right\},$$

τότε

$$\vartheta\in I_a(x)\Leftrightarrow\pi(\vartheta\,|\,x)\ge\gamma_a \text{ και } \vartheta'\notin I_a(x)\Leftrightarrow\pi(\vartheta'\,|\,x)<\gamma_a \text{ έτσι } \pi(\vartheta\,|\,x)>\pi(\vartheta'\,|\,x).$$

Δηλαδή **τα HPD σύνολα δίνουν** $P\left\{\vartheta\in I_a(x)\,|\,x\right\}=1-a$ **αλλά ταυτόχρονα έχουν και ελάχιστο μήκος**. **Επίσης τα σύνολα HPD, μπορεί να μην είναι διαστήματα αλλά ένωση διαστημάτων όταν η** $\pi(\vartheta\,|\,x)$ **είναι πολυκόρυφη.**

Η παρακάτω συνάρτηση της R υπολογίζει το 90% HPDI για ένα Binomial – Beta μοντέλο με $n=70$, διωνυμική παρατήρηση $x=34$, prior $\pi(\vartheta)=Be(\vartheta|4.4,6.6)$ και posterior $\pi(\vartheta|x=34)=Be(\vartheta|38.4,42.6)$




```
HPDIbeta<-function(p=4.4, q=6.6, n=70, y=34, a=0.05){
  library(TeachingDemos)
  curve(dbeta(x, shape1=p+y, shape2=q+n-y), xlim=c(0,1), ylim=c(0,7),
        xlab="theta", ylab="pi(theta|x=34)", , col='black')
  curve(dbeta(x, shape1=p, shape2=q), xlim=c(0,1), add=TRUE)
  xpts<-hpd(qbeta, shape1=p+y, shape2=q+n-y, conf=1-a)
  ypts<-dbeta(xpts, shape1=p+y, shape2=q+n-y)

  lines(xpts, ypts, lty=5, col='red')
  lines(c(xpts[1],xpts[1]), c(0, ypts[1]), lty=5, col='red')
  lines(c(xpts[2],xpts[2]), c(0, ypts[2]), lty=5, col='red')

  return(xpts)
}

x<-HPDIbeta(a=0.1)
> x
[1] 0.3832362 0.5647785
```

<u>Πρόταση</u>: Όταν η prior είναι μίξη κατανομών, τότε και η posterior είναι μίξη κατανομών.

1. Έστω ότι η prior αποτελεί διακριτή μίξη

$$I \sim \sum_{i=1}^{m} p_{i,0}\, \delta_i(\,\cdot\,)$$

$$\vartheta \mid I = i \sim \pi(\,\cdot \mid I = i) = \pi_i(\,\cdot\,)$$

τότε

$$\pi(\vartheta)=\sum_{i=1}^{m}\pi(\vartheta,I=i)=\sum_{i=1}^{m}\pi(I=i)\pi(\vartheta\,|\,I=i)=\sum_{i=1}^{m}p_{i,0}\pi_i(\vartheta).$$

Η posterior θα είναι

$$\pi(\vartheta\,|\,x)\propto\left\{\sum_{i=1}^{m}p_{i,0}\pi_i(\vartheta)\right\}\pi(x\,|\,\vartheta)=\sum_{i=1}^{m}p_{i,0}\pi_i(\vartheta)\pi(x\,|\,\vartheta)$$

$$\Rightarrow C^{-1}=\sum_{i=1}^{m}p_{i,0}\int_{\Theta}\pi(x\,|\,\vartheta)\pi_i(\vartheta)d\vartheta=\sum_{i=1}^{m}p_{i,0}\int_{\Theta}\pi_i(x,\vartheta)d\vartheta=\sum_{i=1}^{m}p_{i,0}\pi_i(x),$$

εναλλακτικά

$$\pi(\vartheta\,|\,x)\propto\pi(\vartheta,x)=\sum_{i=1}^{m}\pi(\vartheta,x,I=i)=\sum_{i=1}^{m}\pi(I=i)\underbrace{\pi(\vartheta\,|\,I=i)\pi(x\,|\,\vartheta,I=i)}_{\pi(\vartheta,x\,|\,I=i)}$$

$$=\sum_{i=1}^{m}\pi(I=i)\pi(x\,|\,I=i)\underbrace{\frac{\pi(\vartheta\,|\,I=i)\pi(x\,|\,\vartheta,I=i)}{\pi(x\,|\,I=i)}}_{\pi_i(\vartheta|x)}$$

$$\Rightarrow C^{-1}=\sum_{i=1}^{m}\underbrace{\pi(I=i)}_{p_{i,0}}\underbrace{\pi(x\,|\,I=i)}_{\pi_i(x)}\underbrace{\int_{\Theta}\pi_i(\vartheta\,|\,x)d\vartheta}_{1}=\sum_{i=1}^{m}p_{i,0}\pi_i(x),$$

με αποτέλεσμα η posterior να είναι και αυτή διακριτή μίξη:

$$\pi(\vartheta\,|\,x)=\sum_{i=1}^{m}\left\{\frac{p_{i,0}}{\sum_{j=1}^{m}p_{j,0}\pi_j(x)}\right\}\pi_i(\vartheta)\pi(x\,|\,\vartheta)=\sum_{i=1}^{m}\left\{\frac{p_{i,0}\pi_i(x)}{\sum_{j=1}^{m}p_{j,0}\pi_j(x)}\right\}\frac{\pi_i(\vartheta)\pi(x\,|\,\vartheta)}{\pi_i(x)}$$

$$=\sum_{i=1}^{m}\left\{\frac{p_{i,0}\pi_i(x)}{\sum_{j=1}^{m}p_{j,0}\pi_j(x)}\right\}\frac{\pi_i(\vartheta)\pi(x\,|\,\vartheta)}{\int_{\Theta}\pi_i(\vartheta)\pi(x\,|\,\vartheta)d\vartheta}=\sum_{i=1}^{m}p_{i,n}\pi_i(\vartheta\,|\,x),$$

όπου $p_{i,n}=\dfrac{p_{i,0}\,\pi_i(x)}{\sum_{j=1}^{m}p_{j,0}\,\pi_j(x)}$ και $\pi_i(\vartheta\,|\,x)=\dfrac{\pi_i(\vartheta)\pi(x\,|\,\vartheta)}{\pi_i(x)}$ για $1\le i\le m$.

2. Εάν η prior αποτελεί συνεχή μίξη θα έχουμε:

$\varphi \sim \pi(\cdot)$

$\vartheta | \varphi \sim \pi(\cdot | \varphi) = \pi_{\varphi}(\cdot)$

τότε

$$\pi(\vartheta) = \int_{\Phi} \pi(\vartheta, \varphi) d\varphi = \int_{\Phi} \pi(\vartheta | \varphi) \pi(\varphi) d\varphi ,$$

και

$$\pi(\vartheta | x) \propto \pi(\vartheta, x) = \int_{\Phi} \pi(\vartheta, \varphi, x) d\varphi = \int_{\Phi} \pi(\varphi) \pi(\vartheta | \varphi) \pi(x | \vartheta, \varphi) d\varphi$$

$$= \int_{\Phi} \pi(\varphi) \pi(x | \varphi) \underbrace{\frac{\pi(\vartheta | \varphi) \pi(x | \vartheta, \varphi)}{\pi(x | \varphi)}}_{\pi(\vartheta | x, \varphi)} d\varphi$$

$$\Rightarrow C^{-1} = \int_{\Phi} \pi(\varphi) \pi(x | \varphi) \int_{\Theta} \pi(\vartheta | x, \varphi) d\vartheta d\varphi = \int_{\Phi} \pi(\varphi) \pi(x | \varphi) d\varphi$$

έτσι η posterior γίνεται

$$\pi(\vartheta | x) = \int_{\Phi} \underbrace{\frac{\pi(\varphi) \pi(x | \varphi)}{\int_{\Phi} \pi(\varphi') \pi(x | \varphi') d\varphi'}}_{\pi(\varphi | x)} \pi(\vartheta | x, \varphi) d\varphi = \int_{\Phi} \pi(\vartheta | x, \varphi) \pi(\varphi | x) d\varphi .$$

Διακριτή μίξη beta κατανομών (beta – mixtures). Όταν οι αρχικές μας πεποιθήσεις είναι πολύπλοκες και τα δεδομένα μας είναι δοκιμές Bernoulli (είτε γεωμετρικές παρατηρήσεις), με προς εκτίμηση πιθανότητα επιτυχίας ϑ η κατάλληλη prior είναι μια διακριτή μίξη κατανομών beta. Για παράδειγμα για μίξη με δύο beta κατανομές[4]:

$$\pi(\vartheta) = \tau_0 Be(\vartheta | p_{1,0}, q_{1,0}) + (1 - \tau_0) Be(\vartheta | p_{2,0}, q_{2,0}), \ 0 < \tau_0 < 1 ,$$

έχουμε πέντε υπερπαραμέτρους τ_0 , $(p_{1,0}, q_{1,0})$ και $(p_{2,0}, q_{2,0})$, και το μοντέλο μας γίνεται πιο αποτελεσματικό (flexible) εφόσον μπορεί να προσαρμοσθεί σε πιο γενικές

[4] A mixture with two beta components.

αρχικές πεποιθήσεις. Θα δείξουμε ότι εάν $x|\vartheta \sim Bin(n,\vartheta)$ με $x = n\overline{x} = \sum_{i=1}^{n} x_i$ και $x_i|\vartheta \sim Bin(1,\vartheta),\, i=1,\ldots,n$, τότε η posterior δίνεται από την beta – μίξη:

$$\pi(\vartheta|x) = \tau_n Be(\vartheta|p_{1,n},q_{1,n}) + (1-\tau_n)Be(\vartheta|p_{2,n},q_{2,n}).$$

Πράγματι

$$\pi(\vartheta|x) \propto \left\{\tau_0 Be(\vartheta|p_{1,0},q_{1,0}) + (1-\tau_0)Be(\vartheta|p_{2,0},q_{2,0})\right\} \times \vartheta^x (1-\vartheta)^{n-x}$$

$$= \frac{\tau_0}{B_{1,0}}\vartheta^{p_{1,0}+x-1}(1-\vartheta)^{q_{1,0}+n-x-1} + \frac{1-\tau_0}{B_{2,0}}\vartheta^{p_{2,0}+x-1}(1-\vartheta)^{q_{2,0}+n-x-1},$$

όπου $B_{i,0} = \dfrac{\Gamma(p_{i,0})\Gamma(q_{i,0})}{\Gamma(p_{i,0}+q_{i,0})} = \int\limits_{\vartheta=0}^{1} \vartheta^{p_{i,0}-1}(1-\vartheta)^{q_{i,0}-1}\, d\vartheta,\, i=1,2$ το ολοκλήρωμα beta. Έστω $C>0$, τέτοιο ώστε:

$$\pi(\vartheta|x) = C\left\{\frac{\tau_0}{B_{1,0}}\vartheta^{p_{1,0}+x-1}(1-\vartheta)^{q_{1,0}+n-x-1} + \frac{1-\tau_0}{B_{2,0}}B_{2,0}\vartheta^{p_{2,0}+x-1}(1-\vartheta)^{q_{2,0}+n-x-1}\right\}.$$

Ολοκληρώνοντας για $\vartheta \in (0,1)$ παίρνουμε

$$C^{-1} = \frac{\tau_0}{B_{1,0}}\int\limits_{\vartheta=0}^{1}\vartheta^{p_{1,0}+x-1}(1-\vartheta)^{q_{1,0}+n-x-1}\, d\vartheta + \frac{1-\tau_0}{B_{2,0}}\int\limits_{\vartheta=0}^{1}\vartheta^{p_{2,0}+x-1}(1-\vartheta)^{q_{2,0}+n-x-1}\, d\vartheta.$$

Θέτοντας $B_{i,n} = \dfrac{\Gamma(p_{i,0}+x)\Gamma(q_{i,0}+n-x)}{\Gamma(p_{i,0}+q_{i,0}+n)}$ παίρνουμε $C^{-1} = \tau_0\dfrac{B_{1,n}}{B_{1,0}} + (1-\tau_0)\dfrac{B_{2,n}}{B_{2,0}}$ και

$$\pi(\vartheta|x) = C\tau_0\frac{B_{1,n}}{B_{1,0}}Be(\vartheta|p_{1,0}+x,q_{1,0}+n-x) + C(1-\tau_0)\frac{B_{2,n}}{B_{2,0}}Be(\vartheta|p_{2,0}+x,q_{2,0}+n-x),$$

όπου ισχύει ότι $C\tau_0\dfrac{B_{1,n}}{B_{1,0}} + C(1-\tau_0)\dfrac{B_{2,n}}{B_{2,0}} = 1$. Θέτοντας $\tau_n = C\tau_0\dfrac{B_{1,n}}{B_{1,0}}$, $p_{i,n} = p_{i,0}+x$ και $q_{i,n} = q_{i,0}+n-x$, έχουμε ότι:

$$\pi(\vartheta|x) = \tau_n Be(\vartheta|p_{1,n},q_{1,n}) + (1-\tau_n)Be(\vartheta|p_{2,n},q_{2,n}),\ 0<\tau_n<1.$$

<u>Άσκηση</u>

Εάν $x\mid\vartheta \sim Bin(n,\vartheta)$ με $x = n\overline{x} = \sum_{i=1}^{n} x_i$ και $x_i\mid\vartheta \sim Bin(1,\vartheta),\, i=1,\ldots,n$, να βρεθεί η posterior και οι εκτίμηση κατά Bayes του ϑ, κάτω από τετραγωνική συνάρτηση απώλειας που αντιστοιχεί στην prior $\pi(\vartheta) \propto \sum_{j=1}^{m} \vartheta^{p_{j,0}-1}(1-\vartheta)^{q_{j,0}-1}$.

<u>Άσκηση</u>

1. Εάν $\pi(\vartheta) \propto \kappa + \lambda\vartheta^{a-1}(1-\vartheta)^{b-1}$ για $0<\vartheta<1$ και $\kappa>0$, $\lambda>0$ δείξτε ότι $\pi(\vartheta) = \rho\,\mathcal{U}(\vartheta\mid 0,1) + (1-\rho)Be(\vartheta\mid a,b)$ με $\rho = \dfrac{\kappa}{\kappa+\lambda B(a,b)}$.

2. Δίνεται ότι $x\mid\vartheta \sim Bin(n,\vartheta)$ με $x=1$ και $n=3$ να βρεθεί η posterior και οι εκτίμηση κατά Bayes του ϑ, κάτω από τετραγωνική συνάρτηση απώλειας, που αντιστοιχεί στην prior $\pi(\vartheta) \propto 1+\vartheta(1-\vartheta)$.

1. Ολοκληρώνοντας την $\pi(\vartheta) = C\left\{\kappa + \lambda\vartheta^{a-1}(1-\vartheta)^{b-1}\right\}1(0<\vartheta<1)$ έχουμε ότι $C^{-1} = \kappa + \lambda B(a,b)$ που δίνει

$$\pi(\vartheta) = \frac{\kappa}{\kappa+\lambda B(a,b)}1(0<\vartheta<1) + \frac{\lambda}{\kappa+\lambda B(a,b)}\vartheta^{a-1}(1-\vartheta)^{b-1}1(0<\vartheta<1)$$

$$= \frac{\kappa}{\kappa+\lambda B(a,b)}\mathcal{U}(\vartheta\mid 0,1) + \frac{\lambda B(a,b)}{\kappa+\lambda B(a,b)}Be(\vartheta\mid a,b)$$

2. $L(\vartheta; x=1) = \pi(x=1\mid\vartheta) \overset{\vartheta}{\propto} \vartheta(1-\vartheta)^2$

$$\pi(\vartheta\mid x=1) \propto \{1+\vartheta(1-\vartheta)\}\times\vartheta(1-\vartheta)^2 = \vartheta(1-\vartheta)^2 + \vartheta^2(1-\vartheta)^3$$

Η σταθερά κανονικοποίησης C της posterior είναι $C^{-1} = B(2,3)+B(3,4)$ και έτσι

$$\pi(\vartheta\mid x=1) = \frac{B(2,3)}{B(2,3)+B(3,4)}Be(\vartheta\mid 2,3) + \frac{B(3,4)}{B(2,3)+B(3,4)}Be(\vartheta\mid 3,4).$$

Η εκτίμηση κατά Bayes του ϑ, κάτω από τετραγωνική συνάρτηση απώλειας είναι

$$\mathbb{E}(\vartheta \mid x=1) = \frac{B(2,3)}{B(2,3)+B(3,4)}\left(\frac{2}{5}\right) + \frac{B(3,4)}{B(2,3)+B(3,4)}\left(\frac{3}{7}\right)$$

Έχουμε ότι $B(2,3) = \frac{1}{12}$, $B(3,4) = \frac{1}{15}$ που δίνουν $\mathbb{E}(\vartheta \mid x=1) = \frac{13}{31}$.

Άλλος τρόπος με τον οποίο, στο binomial – beta μοντέλο, μπορεί να προκύψει σαν posterior beta – μίξη, είναι η πληροφορία να για το διωνυμικό πείραμα:

$x \mid \vartheta \sim Bin(n,\vartheta)$ με $x = n\overline{x} = \sum_{i=1}^{n} x_i$ και $x_i \mid \vartheta \sim Bin(1,\vartheta),\, i = 1,\ldots,n$,

να δίνεται με την μορφή του ενδεχομένου

$Data = \{x = m_1\} \cup \cdots \cup \{x = m_r\}$ με $0 \le m_i \le n$.

Για παράδειγμα, έστω ότι δίνεται η πληροφορία ότι
$Data = \{x \le m\} = \{x = 0\} \cup \cdots \cup \{x = m\}$, τότε η πιθανοφάνεια δίνεται από την εξίσωση

$$L(\vartheta; Data) = \pi(x \le m \mid \vartheta) = \sum_{x=0}^{m} \binom{n}{x} \vartheta^{x} (1-\vartheta)^{n-x}$$

Θέτοντας σαν prior $\pi(\vartheta) = Be(\vartheta \mid a,b)$, η posterior είναι:

$$\pi(\vartheta|x) \propto \vartheta^{a-1}(1-\vartheta)^{b-1} \times \sum_{x=0}^{m} \binom{n}{x} \vartheta^{x}(1-\vartheta)^{n-x} = \sum_{x=0}^{m} \binom{n}{x} \vartheta^{x+a-1}(1-\vartheta)^{b+n-x-1},$$

με σταθερά κανονικοποίησης

$$C^{-1} = \sum_{x=0}^{m} \binom{n}{x} \int_{\vartheta=0}^{1} \vartheta^{a+x-1}(1-\vartheta)^{b+n-x-1}\, d\vartheta = \sum_{x=0}^{m} \binom{n}{x} B(a+x, b+n-x),$$

που δίνει

$$\pi(\vartheta|x) = \sum_{x=0}^{m} C \binom{n}{x} \vartheta^{a+x-1}(1-\vartheta)^{b+n-x-1}$$

$$= \sum_{x=0}^{m} C \binom{n}{x} B(a+x, b+n-x) \frac{\vartheta^{a+x-1}(1-\vartheta)^{b+n-x-1}}{B(a+x, b+n-x)}$$

$$= \sum_{x=0}^{m} p(x \mid n,a,b) Be(\vartheta \mid a+x, b+n-x)$$

όπου

$$p(x \mid n,a,b) = \frac{\binom{n}{x} B(a+x, b+n-x)}{\sum_{y=0}^{m} \binom{n}{y} B(a+y, b+n-y)} \text{ και } \sum_{x=0}^{m} p(x \mid n,a,b) = 1 .$$

Εμφανώς η εκτίμηση κάτω από τετραγωνική συνάρτηση απώλειας θα είναι

$$\mathbb{E}(\vartheta \mid x) = \sum_{x=0}^{m} p(x \mid n,a,b) \int_{\vartheta=0}^{1} \vartheta Be(\vartheta \mid a+x, b+n-x) d\vartheta = \sum_{x=0}^{m} p(x \mid n,a,b) \frac{a+x}{a+b+n} .$$

<u>Άσκηση</u>

Έστω ότι πραγματοποιείται διωνυμικό πείραμα με $n=10$ δοκιμές Bernoulli. Αλλά για κάποιο λόγο, η τελική πληροφορία από τη διεξαγωγή του πειράματος που φτάνει σε εμάς, είναι η πραγματοποίηση του ενδεχομένου $\{x \le 2\}$. Ποία η εκτίμηση της πιθανότητας επιτυχίας ϑ ως προς τετραγωνική συνάρτηση απώλειας εάν οι αρχικές μας πεποιθήσεις συνοψίζονται στην prior $\pi(\vartheta) = Be(\vartheta \mid 4,4)$;

Η πιθανοφάνεια δίνεται από τη σχέση

$$L(\vartheta; x \le 2) = \pi(x \le 2 \mid \vartheta) = \sum_{x=0}^{2} \binom{10}{x} \vartheta^{x} (1-\vartheta)^{10-x}$$

$$\pi(\vartheta \mid x \le 2) \propto \vartheta^{3} (1-\vartheta)^{3} \times \sum_{x=0}^{2} \binom{10}{x} \vartheta^{x} (1-\vartheta)^{10-x} = \sum_{x=0}^{2} \binom{10}{x} \vartheta^{x+3} (1-\vartheta)^{13-x}$$

$$C^{-1} = \sum_{x=0}^{2} \binom{10}{x} B(x+4, 14-x) = B(4,14) + 10B(5,13) + 45B(6,12)$$

$$= \frac{1}{\Gamma(18)} \left[\Gamma(4)\Gamma(14) + 10\Gamma(5)\Gamma(13) + 45\Gamma(6)\Gamma(12) \right]$$

$$=\frac{\Gamma(4)\Gamma(12)}{\Gamma(18)}\left[13\times12+10\times4\times12+45\times5\times4\right]=1536\times\frac{\Gamma(4)\Gamma(12)}{\Gamma(18)}$$

Η posterior είναι

$$\pi(\vartheta|x\le 2)=\sum_{x=0}^{2}C\binom{10}{x}B(4+x,14-x)Be(\vartheta\,|\,4+x,14-x)$$

Η εκτίμηση της πιθανότητας επιτυχίας ϑ είναι

$$\mathbb{E}(\vartheta\,|\,x\le 2)=C\sum_{x=0}^{m}\binom{10}{x}B(4+x,14-x)\frac{4+x}{18}$$

$$=C\left[B(4,14)\frac{4}{18}+10B(5,13)\frac{5}{18}+45B(6,12)\frac{6}{18}\right]$$

$$=\frac{C}{18\Gamma(18)}\left[\Gamma(4)\Gamma(14)\times4+10\times\Gamma(5)\Gamma(13)\times5+45\times\Gamma(6)\Gamma(12)\times6\right]$$

$$=\frac{\Gamma(18)}{1536\times\Gamma(4)\Gamma(12)}\frac{4\times\Gamma(4)\Gamma(12)}{18\times\Gamma(18)}\left[13\times12+10\times12\times5+45\times5\times6\right]=\frac{2106}{1536}\frac{4}{18}=\frac{39}{128}.$$

<u>Άσκηση</u>

Έστω ότι πραγματοποιείται διωνυμικό πείραμα με n δοκιμές Bernoulli και έχουμε n παρατηρήσεις $x_i\,|\,\vartheta \overset{iid}{\sim} Bin(1,\vartheta),\ i=1,\dots,n$. Στη συνέχεια πραγματοποιείται και δεύτερο διωνυμικό πείραμα όμως με πιθανότητα επιτυχίας ϑ/k, δηλαδή μας δίνεται παρατήρηση $z=x_{n+1}$ τέτοια ώστε $z\,|\,\vartheta\sim Bin(1,\vartheta/k)$ και $[z\,|\,\vartheta]$ ανεξάρτητη των $[x_i\,|\,\vartheta]$ για $i=1,\dots,n$. Να βρεθούν οι εκτιμήσεις της πιθανότητας επιτυχίας ϑ ως προς τετραγωνική συνάρτηση απώλειας $\mathbb{E}(\vartheta\,|\,x_1,\dots,x_n,z=0)$ και $\mathbb{E}(\vartheta\,|\,x_1,\dots,x_n,z=1)$, υποθέτοντας ότι $\pi(\vartheta)=Be(\vartheta\,|\,a,b)$. Να γίνει αριθμητική εφαρμογή για $n=2$, $n\overline{x}=1$, και $k=2$.

Για την πραγματοποίηση $\{z=0\}$ έχουμε

$$\pi\left(z=0\mid\vartheta\right)=1-\frac{\vartheta}{k}=\frac{1}{k}\left[\left(k-1\right)+\left(1-\vartheta\right)\right],$$

και έτσι

$$\pi\left(x_1,\ldots,x_n,z=0\mid\vartheta\right)=\pi\left(x_1,\ldots,x_n\mid\vartheta\right)\pi\left(z=0\mid\vartheta\right)$$

$$=\prod_{i=1}^{n}\vartheta^{x_i}\left(1-\vartheta\right)^{1-x_i}\times\frac{1}{k}\left[\left(k-1\right)+\left(1-\vartheta\right)\right]$$

$$\overset{\vartheta}{\propto}\left(k-1\right)\vartheta^{n\bar{x}}\left(1-\vartheta\right)^{n-n\bar{x}}+\vartheta^{n\bar{x}}\left(1-\vartheta\right)^{n-n\bar{x}+1}.$$

Η posterior γίνεται

$$\pi\left(\vartheta\mid x_1,\ldots,x_n,z=0\right)\propto\vartheta^{a-1}\left(1-\vartheta\right)^{b-1}\times\left[\left(k-1\right)\vartheta^{n\bar{x}}\left(1-\vartheta\right)^{n-n\bar{x}}+\vartheta^{n\bar{x}}\left(1-\vartheta\right)^{n-n\bar{x}+1}\right]$$

$$=\left(k-1\right)\vartheta^{a+n\bar{x}-1}\left(1-\vartheta\right)^{b+n-n\bar{x}-1}+\vartheta^{a+n\bar{x}-1}\left(1-\vartheta\right)^{b+n-n\bar{x}}.$$

Με σταθερά κανονικοποίησης C

$$C^{-1}=\left(k-1\right)B\left(a+n\bar{x},b+n-n\bar{x}\right)+B\left(a+n\bar{x},b+n-n\bar{x}+1\right).$$

Η posterior αναπαρίσταται σαν mixture από betas

$$\pi\left(\vartheta\mid x_1,\ldots,x_n,z=0\right)=C\left(k-1\right)B\left(a+n\bar{x},b+n-n\bar{x}\right)Be\left(\vartheta\mid a+n\bar{x},b+n-n\bar{x}\right)$$

$$+CB\left(a+n\bar{x},b+n-n\bar{x}+1\right)Be\left(\vartheta\mid a+n\bar{x},b+n-n\bar{x}+1\right),$$

και μέση τιμή

$$\mathbb{E}\left(\vartheta\mid x_1,\ldots,x_n,z=0\right)=C\left(k-1\right)B\left(a+n\bar{x},b+n-n\bar{x}\right)\left(\frac{a+n\bar{x}}{a+b+n}\right)$$

$$+C\,B\left(a+n\bar{x},b+n-n\bar{x}+1\right)\left(\frac{a+n\bar{x}}{a+b+n+1}\right),$$

ενώ παρατηρούμε ότι:

$$B\left(a+n\bar{x},b+n-n\bar{x}+1\right)=B\left(a+n\bar{x},b+n-n\bar{x}\right)\frac{b+n-n\bar{x}}{a+b+n}.$$

Για $n=2$, $n\bar{x}=1$, και $k=2$ παίρνουμε:

$$\pi\left(\vartheta \mid x_1,\dots,x_n,z=0\right)=\frac{a+b+2}{a+2b+3}Be\left(\vartheta \mid a+1,b+1\right)+\frac{b+1}{a+2b+3}Be\left(\vartheta \mid a+1,b+2\right),$$

$$\mathbb{E}\left(\vartheta \mid x_1,\dots,x_n,z=0\right)=\frac{a+b+2}{a+2b+3}\frac{a+1}{a+b+2}+\frac{b+1}{a+2b+3}\frac{a+1}{a+b+3}=\frac{a+1}{a+2b+3}\left(1+\frac{b+1}{a+b+3}\right).$$

Για την πραγματοποίηση $\{z=1\}$ έχουμε $\pi\left(z=1 \mid \vartheta\right)=\frac{\vartheta}{k}$, που δίνει

$$\pi\left(x_1,\dots,x_n,z=1 \mid \vartheta\right)=\pi\left(x_1,\dots,x_n \mid \vartheta\right)\pi\left(z=1 \mid \vartheta\right)=\prod_{i=1}^{n}\vartheta^{x_i}\left(1-\vartheta\right)^{1-x_i}\times\frac{\vartheta}{k}\underset{\propto}{\vartheta}\ \vartheta^{n\bar{x}+1}\left(1-\vartheta\right)^{n-n\bar{x}}.$$

Η posterior γίνεται

$$\pi\left(\vartheta \mid x_1,\dots,x_n,z=1\right)\propto\vartheta^{a-1}\left(1-\vartheta\right)^{b-1}\times\vartheta^{n\bar{x}+1}\left(1-\vartheta\right)^{n-n\bar{x}}\propto Be\left(\vartheta \mid a+n\bar{x}+1,b+n-n\bar{x}\right),$$

με $\mathbb{E}\left(\vartheta \mid x_1,\dots,x_n,z=1\right)=\frac{a+n\bar{x}+1}{a+b+n+1}$.

Στην ειδική περίπτωση $n=2$, $n\bar{x}=1$, και $k=2$ παίρνουμε:

$\pi\left(\vartheta \mid x_1,\dots,x_n,z=1\right)=Be\left(\vartheta \mid a+2,b+1\right)$ με $\mathbb{E}\left(\vartheta \mid x_1,\dots,x_n,z=1\right)=\frac{a+2}{a+b+3}$.

<u>Άσκηση</u>

Έστω ότι πραγματοποιούντα δύο δοκιμές Bernoulli $x \mid \vartheta \sim Bin(1,\vartheta)$, και $y \mid \vartheta \sim Bin(1,\vartheta/3)$, με $[x \mid \vartheta]$ και $[y \mid \vartheta]$ ανεξάρτητες τυχαίες μεταβλητές και παρατηρούμε ότι $x=1$ και $y=0$. Ποία η εκτίμηση του ϑ κάτω από τετραγωνική συνάρτηση απώλειας και κάτω από απουσία αρχικών πεποιθήσεων? Με ποία πιθανότητα στο μέλλον θα παρατηρήσουμε $z=1$, εάν $z \mid \vartheta \sim Bin(1,\vartheta)$?

$$\pi\left(x=1,y=0 \mid \vartheta\right)=\pi\left(x=1 \mid \vartheta\right)\pi\left(y=0 \mid \vartheta\right)=\vartheta\times\left(1-\frac{\vartheta}{3}\right)$$

$$=\vartheta\times\frac{1}{3}\left[2+\left(1-\vartheta\right)\right]\propto 2\vartheta+\vartheta\left(1-\vartheta\right)$$

$$\pi\left(\vartheta \mid x=1,y=0\right)\propto\mathcal{U}\left(\vartheta|0,1\right)\times\left\{2\vartheta+\vartheta\left(1-\vartheta\right)\right\}=2\vartheta+\vartheta\left(1-\vartheta\right)$$

Η σταθερά κανονικοποίησης C είναι

$$\pi\left(\vartheta \mid x=1, y=0\right)=C\left\{2\vartheta+\vartheta\left(1-\vartheta\right)\right\} \Rightarrow 1=C\left\{2B\left(2,1\right)+B\left(2,2\right)\right\} \Rightarrow C=\frac{6}{7}$$

$$\pi\left(\vartheta \mid x=1, y=0\right)=\frac{6}{7}\left\{2\vartheta+\vartheta\left(1-\vartheta\right)\right\}$$

$$=\frac{6}{7}C\left\{2B\left(2,1\right)B\left(\vartheta \mid 2,1\right)+B\left(2,2\right)B\left(\vartheta \mid 2,2\right)\right\}$$

$$=\frac{6}{7}B\left(\vartheta \mid 2,1\right)+\frac{1}{7}B\left(\vartheta \mid 2,2\right)$$

$$\mathbb{E}\left(\vartheta \mid x=1, y=0\right)=\frac{6}{7}\int_0^1 \vartheta B\left(\vartheta \mid 2,1\right)d\vartheta+\frac{1}{7}\int_0^1 \vartheta B\left(\vartheta \mid 2,2\right)d\vartheta$$

$$=\frac{6}{7}\frac{B\left(3,1\right)}{B\left(2,1\right)}+\frac{1}{7}\frac{B\left(3,2\right)}{B\left(2,2\right)}=\frac{9}{14}$$

Το posterior predictive είναι:

$$\pi\left(z \mid x=1, y=0\right)=\int_0^1 \pi\left(z \mid \vartheta\right)\pi\left(\vartheta \mid x=1, y=0\right)d\vartheta=\frac{6}{7}\int_0^1 \vartheta^z\left(1-\vartheta\right)^{1-z}\left\{2\vartheta+\vartheta\left(1-\vartheta\right)\right\}d\vartheta$$

$$=\frac{6}{7}\left\{2B\left(z+2,2-z\right)+B\left(z+2,3-z\right)\right\},\ z\in\left\{0,1\right\}$$

Που δίνει

$$\pi\left(z=1 \mid x=1, y=0\right)=\frac{6}{7}\left\{2B\left(3,1\right)+B\left(3,2\right)\right\}=\frac{9}{14}$$