Επάρκεια στην κλασσική Στατιστική: Η επάρκεια είναι μια ιδιότητα που έχει μια στατιστική (δειγματική) συνάρτηση $t=t(x)$, με $x=(x_1,\dots,x_n)$ ως προς μια παράμετρο ϑ. Πιο συγκεκριμένα η $t(x)$ είναι επαρκής όταν $\pi(x|t,\vartheta)=\pi(x|t)$. Με άλλα λόγια **η πληροφορία που μεταφέρει η ϑ για το δείγμα x είναι η ίδια με την πληροφορία που μεταφέρει η $t(x)$.**

Επάρκεια κατά Bayes: Κάτω από την οπτική της στατιστικής κατά Bayes θα πρέπει να ισχύει η ισότητα για την posterior

$$[\vartheta|x] \overset{d}{=} [\vartheta|t(x)].$$

Δηλαδή η παράμετρος ϑ, αλληλεπιδρά με το x μέσα στην posterior, μόνο μέσο της $t(x)$. Πράγματι εάν $\pi(x|t,\vartheta)=\pi(x|t)$ τότε πολλαπλασιάζοντας και τα δύο μέλη με $\pi(t,\vartheta)$ παίρνουμε

$$\pi(x,t,\vartheta)=\pi(t,\vartheta)\pi(x|t) \Leftrightarrow \pi(x,t,\vartheta)=\pi(t)\pi(\vartheta|t)\pi(x|t)$$

$$\Leftrightarrow \pi(\vartheta|x,t)\pi(x,t)=\pi(\vartheta|t)\pi(x,t) \Leftrightarrow \pi(\vartheta|x,t)=\pi(\vartheta|t).$$

Το $t(x)$ γενικά μεταφέρει λιγότερη πληροφορία από ότι το x, δηλαδή $\pi(\vartheta|x,t)=\pi(\vartheta|x)$ και έτσι από την προηγούμενη σχέση έχουμε $\pi(\vartheta|x)=\pi(\vartheta|t)$.

Εναλλακτικά

$$\pi(x|t,\vartheta)=\pi(x|t) \Leftrightarrow \frac{\pi(x)\pi(t,\vartheta|x)}{\pi(t,\vartheta)}=\frac{\pi(x)\pi(t|x)}{\pi(t)} \Leftrightarrow$$

$$\frac{\pi(t,\vartheta|x)}{\pi(t)\pi(\vartheta|t)}=\frac{\pi(t|x)}{\pi(t)} \Leftrightarrow \frac{\pi(t,\vartheta|x)}{\pi(\vartheta|t)}=\pi(t|x) \Leftrightarrow$$

$$\frac{\pi(t|x)\pi(\vartheta|t,x)}{\pi(\vartheta|t)}=\pi(t|x) \Leftrightarrow \frac{\pi(\vartheta|t,x)}{\pi(\vartheta|t)}=1$$

Το θεώρημα των Fisher-Neyman μας λέει ότι η στατιστική $t(x)$ είναι επαρκής εάν και μόνον εάν η συνάρτηση πιθανοφάνειας παραγοντοποιείται στη μορφή

$$\pi(x|\vartheta) = h_{FN}(x)\, g_{FN}(t(x), \vartheta).$$

Δηλαδή η παράμετρος ϑ αλληλεπιδρά μέσα στην πιθανοφάνεια με το x μόνο μέσω της επαρκούς στατιστικής $t(x)$.

Πιθανοφάνεια Bernoulli

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

$$\pi(x|\vartheta) = \prod_{i=1}^{n} \pi(x_i|\vartheta) = \prod_{i=1}^{n} \vartheta^{x_i}(1-\vartheta)^{1-x_i} = \vartheta^{n\bar{x}}(1-\vartheta)^{n(1-\bar{x})}$$

$$h_{FN}(x) = 1,\ g_{FN}(t(x), \vartheta) = \vartheta^{nt(x)}(1-\vartheta)^{n(1-t(x))} \Rightarrow t_{Suff}(x) = \bar{x}.$$

Πιθανοφάνεια Poisson

$[x_i|\vartheta] \overset{iid}{\sim} Po(\cdot|\vartheta),\ 1 \le i \le n$,

$$\pi(x|\vartheta) = \prod_{i=1}^{n} \frac{\vartheta^{x_i} e^{-\vartheta}}{x_i!} = \left\{\prod_{i=1}^{n} x_i!\right\}^{-1} \exp\{n\bar{x}\log(\vartheta) - n\vartheta\}$$

$$h_{FN}(x) = \left(\prod_{i=1}^{n} x_i!\right)^{-1},\ g_{FN}(t(x), \vartheta) = \exp\{n\bar{x}\log(\vartheta) - n\vartheta\} \Rightarrow t_{Suff}(x) = \bar{x}.$$

Πιθανοφάνεια Uniform στο διάστημα $(0, \vartheta)$

$[x_i|\vartheta] \overset{iid}{\sim} \mathcal{U}(\cdot|0, \vartheta),\ 1 \le i \le n,\ \vartheta > 0$

$$\pi(x|\vartheta) = \prod_{i=1}^{n} \mathcal{U}(x_i|0, \vartheta) = \prod_{i=1}^{n} \vartheta^{-1} 1(0 \le x_i \le \vartheta) = \vartheta^{-n} 1(\vartheta \ge x_{(n)})$$

$$h_{FN}(x) = 1,\ g_{FN}(t(x), \vartheta) = \vartheta^{-n} 1(\vartheta \ge x_{(n)}) \Rightarrow t_{Suff}(x) = x_{(n)}$$

όπου $x_{(1)} = \min_{1 \le i \le n} x_i$ και $x_{(n)} = \max_{1 \le i \le n} x_i$ είναι η μικρότερη και μεγαλύτερη διατακτική στατιστική (order statistic) αντιστοίχως.

Εμφανώς $\hat{\vartheta}_{MLE} = x_{(n)}$ διότι $\vartheta^{-n} \downarrow$ και

$$\sup_{\vartheta>0} \pi\left(x|\vartheta\right) = \sup_{\vartheta>0}\left[\vartheta^{-n} 1\left(\vartheta \geq x_{(n)}\right)\right] = \sup_{\vartheta \geq x_{(n)}} \vartheta^{-n} = x_{(n)}^{-n} = \pi\left(x|x_{(n)}\right).$$

<u>Άσκηση </u>Να βρεθούν οι επαρκείς στατιστικές και ο MLE όταν

$x_i|\vartheta \overset{iid}{\sim} Exp\left(\cdot|\vartheta\right)$, $x_i|\vartheta \overset{iid}{\sim} Geo\left(\cdot|\vartheta\right)$, $x|\vartheta \overset{iid}{\sim} Nb\left(\cdot|n,\vartheta\right)$, $x_i|\vartheta \overset{iid}{\sim} Ga\left(\cdot|a,\vartheta\right)$

Πιθανοφάνεια με παρατηρήσεις από την Exponential

$x_i|\vartheta \overset{iid}{\sim} Exp\left(\cdot|\vartheta\right),\ 1 \leq i \leq n$

$$\pi\left(x|\vartheta\right) = \prod_{i=1}^{n} \vartheta e^{-\vartheta x_i} = \vartheta^n e^{-n\vartheta\bar{x}}$$

$$h_{FN}\left(x\right) = 1,\ g_{FN}\left(t\left(x\right),\vartheta\right) = \vartheta^n e^{-n\vartheta\bar{x}} \Rightarrow t\left(x\right) = \bar{x}.$$

Πιθανοφάνεια με παρατηρήσεις από την Geometric

$x_i|\vartheta \overset{iid}{\sim} Geo\left(\cdot|\vartheta\right),\ 1 \leq i \leq n$, $x_i =$ αριθμός των αποτυχιών έως την 1η επιτυχία

$$\pi\left(x|\vartheta\right) = \prod_{i=1}^{n} \vartheta\left(1-\vartheta\right)^{x_i} = \vartheta^n\left(1-\vartheta\right)^{n\bar{x}}$$

$$\Rightarrow h_{FN}\left(x\right) = 1,\ g_{FN}\left(t;\vartheta\right) = \vartheta^n\left(1-\vartheta\right)^{n\bar{x}},\ t = \bar{x}$$

Εκτίμηση παραμέτρου από Poisson data

Υποθέτουμε ότι οι παρατηρήσεις $x_1,\ldots,x_n$ δοθείσας της παραμέτρου ϑ είναι ανεξάρτητες και ακολουθούν την Poisson με παράμετρο ϑ δηλαδή $x_i|\vartheta \overset{iid}{\sim} Po\left(\cdot|\vartheta\right)$. Εάν υποθέσουμε ότι η prior ακολουθεί την Gamma κατανομή $\vartheta \sim Ga\left(\cdot|p,q\right)$ θα έχουμε

$$\pi\left(\vartheta|x\right) \propto \pi\left(\vartheta\right)\pi\left(x|\vartheta\right) = Ga\left(\vartheta|p,q\right)\prod_{i=1}^{n} Po\left(x_i|\vartheta\right)$$

$$\propto \left(\vartheta^{p-1}e^{-q\vartheta}\right)\prod_{i=1}^{n} e^{-\vartheta}\vartheta^{x_i} = \vartheta^{p+n\bar{x}-1}e^{-(n+q)\vartheta}$$

που δίνει $\pi(\vartheta|x) = Ga(\vartheta|\, p + n\overline{x},\, n+q)$

Παρατηρήσεις

- Παρατηρήστε ότι εάν $\wp$ είναι η οικογένεια των $Ga(\,\cdot|\,p,q)$ prior κατανομών και $\Im$ η οικογένεια των $Po(\,\cdot|\vartheta)$ δειγματοληπτικών κατανομών, τότε η posterior ανήκει και αυτή στην οικογένεια $\wp$. Το ίδιο ισχύει και στο beta – binomial υπόδειγμα με prior οικογένεια $\wp$ την οικογένεια των $Be(\,\cdot|\,p,q)$ κατανομών και $\Im$ την οικογένεια των $Bin(\,\cdot|n,\vartheta)$ δειγματοληπτικών κατανομών (για n γνωστό). Σε αυτές τις περιπτώσεις λέμε τότε ότι η οικογένεια $\wp$ είναι συζυγής (conjugate) ως προς την οικογένεια $\Im$.

- Από το αποτέλεσμα $\pi(\vartheta|x) = \pi(\vartheta|\overline{x})$ βλέπουμε ότι μια επαρκής στατιστική είναι η $t(x) = \overline{x}$.

- Ως προς τετραγωνική συνάρτηση απώλειας έχουμε

  $$\widehat{\vartheta}_{Bayes} = \mathbb{E}(\vartheta|x) = \frac{p+n\overline{x}}{q+n} = \left(\frac{q}{q+n}\right)\frac{p}{q} + \left(\frac{n}{q+n}\right)\overline{x} = \nu_n\, \mathbb{E}(\vartheta) + (1-\nu_n)\,\overline{x}$$

  για $\nu_n = \dfrac{q}{q+n} \to 0$ όταν $n \to \infty$

  Δηλαδή έχουμε ότι το Bayes estimator ως προς τετραγωνική συνάρτηση απώλειας είναι κυρτός γραμμικός συνδυασμός του prior mean και του εκτιμητή μεγίστης πιθανοφάνειας

  $\widehat{\vartheta}_{Bayes} = \nu_n\, \mathbb{E}(\vartheta) + (1-\nu_n)\,\widehat{\vartheta}_{MLE} \to \widehat{\vartheta}_{MLE}$ όταν $n \to \infty$.

Το prior predictive για μία Poisson παρατήρηση x είναι **ειδική περίπτωση της Poisson – Gamma κατανομής**

$$PG(x|\,p,q,r) = \int_{\mathbb{R}^+} Po(x|r\vartheta)\,Ga(\vartheta|\,p,q)\,d\vartheta = \int_{\mathbb{R}^+} \frac{e^{-r\vartheta}(r\vartheta)^x}{x!}\frac{q^p}{\Gamma(p)}\vartheta^{p-1}e^{-q\vartheta}d\vartheta$$

$$= \frac{r^x}{x!}\frac{q^p}{\Gamma(p)}\int_{\mathbb{R}^+}\vartheta^{p+x-1}e^{-(q+r)\vartheta}d\vartheta = \frac{q^p}{\Gamma(p)}\frac{r^x}{x!}\frac{\Gamma(p+x)}{(q+r)^{p+x}},\; x \in \mathbb{N}_0$$

Πράγματι $\pi(x)=\int_{\Theta}\pi(x|\vartheta)\Pi(d\vartheta)=\int_{\vartheta>0}Po(x|\vartheta)Ga(\vartheta|p,q)d\vartheta$

$$= PG(x|p,q,1) = \frac{q^p}{\Gamma(p)}\frac{\Gamma(p+x)}{x!(q+1)^{p+x}},\ x\in\mathbb{N}_0 .$$

Όταν $p\in\mathbb{N}$ έχουμε

$$\pi(x)=\frac{\Gamma(p+x)}{\Gamma(p)x!}\left(\frac{q}{q+1}\right)^p\frac{1}{(q+1)^x}=\frac{(p+x-1)!}{(p-1)!\,x!}\left(\frac{q}{q+1}\right)^p\left(1-\frac{q}{q+1}\right)^x$$

$$=\binom{x+p-1}{p-1}\left(\frac{q}{q+1}\right)^p\left(1-\frac{q}{q+1}\right)^x = NB\left(x\,|\,p,\ \frac{q}{q+1}\right),\ x\geq 0 .$$

Δηλαδή, το x είναι ο αριθμός των αποτυχιών, σε αρνητική διωνυμική με p επιτυχίες και πιθανότητα επιτυχίας το $\frac{q}{q+1}$.

Αριθμητικό παράδειγμα

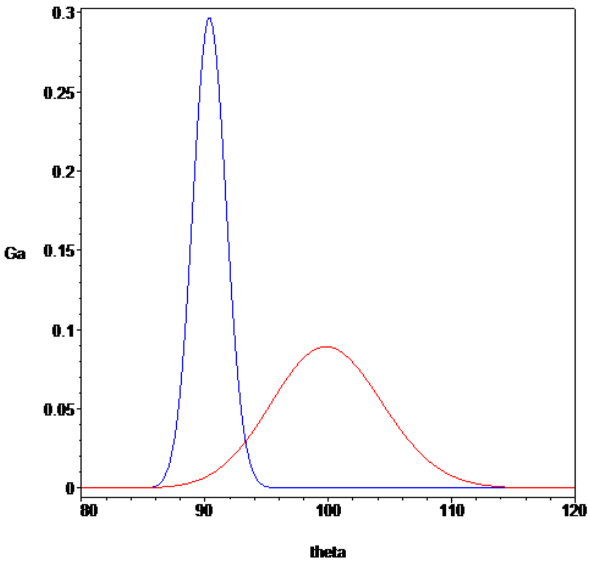
Έστω x_i ο αριθμός πτηνών είδους A στην i φωτογραφία σε μια συγκεκριμένη περιοχή. Μας δίνεται ότι σε συγκεκριμένο data set $x=(x_1,\ldots,x_n)$ για $n=45$ βρέθηκε ότι $n\bar{x}=4019$. Εάν η a-priori πεποίθηση είναι $\mathbb{E}(\vartheta)=100$ και $Var(\vartheta)=20$, να βρεθεί ο εκτιμητής κατά Bayes ως προς τετραγωνική συνάρτηση απώλειας, του αριθμού ϑ των πτηνών είδους A.

Είναι φυσικό να υποθέσουμε ότι $x_i|\vartheta \overset{iid}{\sim} Po(\cdot|\vartheta)$ με Gamma prior Επειδή:

1. Το στήριγμα της άγνωστης κατανομής του ϑ είναι το $\mathbb{R}^+$
2. Η κλάση κατανομών $\vartheta\sim Ga(\cdot|p,q)$ είναι συζυγείς με το παραμετρικό μοντέλο $x_i|\vartheta \overset{iid}{\sim} Po(\cdot|\vartheta)$.
3. Για $p>1$ και $q>0$ η πυκνότητα $Ga(x|p,q)$ είναι μονοκόρυφη.
4. Είναι εύκολο να βρούμε την τιμή των υπερπαραμέτρων (p,q) έτσι ώστε να ενσωματώσουμε στην a-priori κατανομή $\vartheta\sim Ga(\cdot|p,q)$ την πεποίθηση των ειδικών $E(\vartheta)=100$ και $Var(\vartheta)=20$.

Έτσι για $\vartheta \sim Ga(\cdot|p,q)$ θα έχουμε $\mathbb{E}(\vartheta)=p/q=100$ και $Var(\vartheta)=p/q^2=20$. Τότε $p=500$ και $q=5$ ενώ βρίσκουμε ότι τα updates των παραμέτρων του gamma posterior είναι $p+n\overline{x}=4519, n+q=50$ έτσι έχουμε

$$[\vartheta|x] = [\vartheta|n\overline{x}=4019] \sim Ga(\cdot|4519,50)$$



από όπου

$\mathbb{E}(\vartheta)=100$,

$\widehat{\vartheta}_{Bayes}=\mathbb{E}(\vartheta|n\overline{x}=4019)=4519/50=90.380$, $\widehat{\vartheta}_{MLE}=\overline{x}=4019/45=89.311$

$Var(\vartheta|n\overline{x}=4019)=1.808<Var(\vartheta)=20$.

<u>Άσκηση</u>

Εάν $x \sim Ga(a,\vartheta), y \sim Ga(b,\vartheta)$, για $a,b,\vartheta>0$ και x,y ανεξάρτητες τ.μ., δείξτε ότι:

1. $x+y \sim Ga(a+b,\vartheta)$ και ότι ολοκλήρωμα $B(a,b)=\int_0^1 w^{a-1}(1-w)^{b-1}\,dw$ έχει την αναπαράσταση $B(a,b)=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

2. $E\left(X^k\right)=\frac{a_{(k)}}{\vartheta^k}$ και $M_X(t)=\left(\frac{\vartheta}{\vartheta-t}\right)^a$, $|t|<\vartheta$.

1. $\pi(x)=\frac{\vartheta^a}{\Gamma(a)}x^{a-1}e^{-\vartheta x}1(x>0)$, $\pi(y)=\frac{\vartheta^b}{\Gamma(b)}x^{b-1}e^{-\vartheta x}1(y>0)$,

$$T:\begin{Bmatrix} u=x+y \\ v=x \end{Bmatrix} \Leftrightarrow T^{-1}:\begin{Bmatrix} x=v \\ y=u-v \end{Bmatrix} \Rightarrow Jac\left(T^{-1}\right)=\begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix}=-1$$

$$\pi(u,v)=\pi\left(x(u,v),y(u,v)\right)\left|Jac\left(T^{-1}\right)\right|=\pi\left(x(u,v)\right)\pi\left(y(u,v)\right)\left|Jac\left(T^{-1}\right)\right|$$

$$=\frac{\vartheta^a}{\Gamma(a)}v^{a-1}e^{-\vartheta v}1(v>0)\frac{\vartheta^b}{\Gamma(b)}(u-v)^{b-1}e^{-\vartheta(u-v)}1(u-v>0)$$

$$=\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}v^{a-1}(u-v)^{b-1}e^{-\vartheta u}1(u>v>0).$$

Η περιθώρια ως προς u είναι

$$\pi(u)=\int_{\mathbb{R}}\pi(u,v)\,dv=\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}e^{-\vartheta u}\int_{\mathbb{R}}v^{a-1}(u-v)^{b-1}1(u>v>0)\,dv$$

$$=\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}e^{-\vartheta u}\int_{v=0}^{u}v^{a-1}(u-v)^{b-1}\,dv \overset{v=uw}{=} \frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}u^{a+b-1}e^{-\vartheta u}\int_{w=0}^{1}w^{a-1}(1-w)^{b-1}\,dw$$

Επειδή $\pi(u)\propto u^{a+b-1}e^{-\vartheta u}$ θα έχουμε ότι $u=x+y\sim Ga(a+b,\theta)$, τότε

$$1=\int_{\mathbb{R}}\pi(u)\,du=\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}\int_0^1 w^{a-1}(1-w)^{b-1}\,dw\left\{\int_{\mathbb{R}^+}u^{a+b-1}e^{-\vartheta u}du\right\}$$

$$\overset{\tau=\vartheta u}{=}\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}\int_0^1 w^{a-1}(1-w)^{b-1}\,dw\left\{\int_{\mathbb{R}^+}\left(\frac{\tau}{\vartheta}\right)^{a+b-1}e^{-\tau}\left(\frac{d\tau}{\vartheta}\right)\right\}$$

$$=\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}\int_0^1 w^{a-1}(1-w)^{b-1}\,dw\left\{\frac{1}{\vartheta^{a+b}}\int_{\mathbb{R}^+}\tau^{a+b-1}e^{-\tau}d\tau\right\}$$

$$=\frac{\vartheta^{a+b}}{\Gamma(a)\Gamma(b)}\int_0^1 w^{a-1}(1-w)^{b-1}\,dw\left\{\frac{1}{\vartheta^{a+b}}\Gamma(a+b)\right\} \Rightarrow \int_0^1 w^{a-1}(1-w)^{b-1}\,dw=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

2. $E\left(X^k\right)=\int\limits_{\mathbb{R}^+} x^k\left(\frac{\vartheta^a}{\Gamma(a)}x^{a-1}e^{-\vartheta x}dx\right)=\frac{\vartheta^a}{\Gamma(a)}\int\limits_{\mathbb{R}^+}\left(\frac{u}{\vartheta}\right)^{k+a-1}e^{-u}\left(\frac{du}{\vartheta}\right)$

$$=\frac{1}{\vartheta^k\Gamma(a)}\int\limits_{\mathbb{R}^+}u^{k+a-1}e^{-u}du=\frac{\Gamma(a+k)}{\vartheta^k\,\Gamma(a)}=\frac{(a)_k}{\vartheta^k}.$$

$$M_X(t)=E\left(e^{tX}\right)=\sum_{k=0}^{\infty}\frac{t^k}{k!}E\left(X^k\right)=\sum_{k=0}^{\infty}\frac{a_{(k)}}{k!}\left(\frac{t}{\vartheta}\right)^k=\sum_{k=0}^{\infty}\binom{-a}{k}\left(-\frac{t}{\vartheta}\right)^k$$

$$=\left(1-\frac{t}{\vartheta}\right)^{-a}=\left(\frac{\vartheta}{\vartheta-t}\right)^a,\quad \left|\frac{t}{\vartheta}\right|<1\Leftrightarrow|t|<\vartheta.$$

Εναλλακτικά

$$M_X(t)=\int\limits_{\mathbb{R}^+}e^{tx}\left(\frac{\vartheta^a}{\Gamma(a)}x^{a-1}e^{-\vartheta x}dx\right)=\frac{\vartheta^a}{\Gamma(a)}\int\limits_{\mathbb{R}^+}x^{a-1}e^{-(\vartheta-t)x}dx=\frac{\vartheta^a}{\Gamma(a)}\int\limits_{\mathbb{R}^+}\left(\frac{u}{\vartheta-t}\right)^{a-1}e^{-u}\left(\frac{du}{\vartheta-t}\right)$$

$$=\frac{1}{\Gamma(a)}\left(\frac{\vartheta}{\vartheta-t}\right)^a\int\limits_{\mathbb{R}^+}u^{a-1}e^{-u}du=\left(\frac{\vartheta}{\vartheta-t}\right)^a.$$

<u>Άσκηση</u>

Να δειχθεί ότι η κατανομή του αριθμού x των ανεξάρτητων δοκιμών Bernoulli έως την n-οστή επιτυχία και του αριθμού y των αποτυχιών έως την n-οστή επιτυχία είναι:

$$geo(x|\vartheta)=Nb(x|1,\vartheta)=\vartheta(1-\vartheta)^{x-1}\cdot 1(x\ge 1),$$

$$Nb(x|n,\vartheta)=\binom{x-1}{n-1}\vartheta^n(1-\vartheta)^{x-n}\cdot 1(x\ge n),$$

$$Geo(y|\vartheta)=NB(y|1,\vartheta)=\vartheta(1-\vartheta)^y\cdot 1(y\ge 0),$$

$$NB(y|n,\vartheta)=\binom{y+n-1}{n-1}\vartheta^n(1-\vartheta)^y 1(y\ge 0)=\binom{-n}{y}(-1)^y(1-\vartheta)^y 1(y\ge 0).$$

- Δείξτε ότι $\sum_{y=0}^{\infty}\pi(y)=1$

- Βρείτε την πιθανογεννήτρια $G_y(u)$ της $y|\vartheta \sim NB(\cdot|n,\vartheta)$

- Δείξτε ότι $y_i|\vartheta \overset{ind}{\sim} NB(\cdot|n_i,\vartheta)$, $1 \le i \le n \Rightarrow \sum_{i=1}^{n} y_i|\vartheta \sim NB\left(\cdot|\sum_{i=1}^{n} n_i,\vartheta\right)$

Πρώτη παραμετροποίηση: $x =$ ο αριθμός των ανεξάρτητων δοκιμών Bernoulli έως την n-οστή επιτυχία.

$\pi(x) = Nb(x|n,\vartheta)$

$= P\{$ $n-1$ επιτυχίες στις πρώτες $x-1$ δοκιμές, επιτυχία στην x δοκιμή$\}$

$= P\{$ $n-1$ επιτυχίες στις πρώτες $x-1$ δοκιμές$\} \cdot P\{$επιτυχία στην x δοκιμή$\}$

$= Bin(n-1|\, x-1,\vartheta)\ Bin(1|1,\vartheta)$

$$= \left\{\binom{x-1}{n-1}\vartheta^{n-1}(1-\vartheta)^{(x-1)-(n-1)}\right\}\vartheta = \binom{x-1}{n-1}\vartheta^{n}(1-\vartheta)^{x-n}\cdot 1(x \ge n).$$

Δεύτερη παραμετροποίηση: $y = T(x) = x - n =$ *ο αριθμός των αποτυχιών έως την* n-*οστή επιτυχία*.

$$\pi_Y(y) = P\{T(X) = y\} = P\{X = T^{-1}(y)\} = \pi_X\left(T^{-1}(y)\right)$$

$$= \binom{y+n-1}{n-1}\vartheta^{n}(1-\vartheta)^{y}\cdot 1(y \ge 0) = NB(y \mid n,\ \vartheta).$$

$$\vartheta^{-n} = \left[1+\left(-(1-\vartheta)\right)\right]^{-n} = \sum_{y=0}^{\infty}\binom{-n}{y}(-1)^{y}(1-\vartheta)^{y},$$

$$(-1)^{y}\binom{-n}{y} = (-1)^{y}\frac{(-n)(-n-1)\cdots(-n-(y-1))}{y!} = \frac{n(n+1)\cdots(n+(y-1))}{y!}$$

$$\frac{(n-1)!\, n(n+1)\cdots(n+y-1)}{(n-1)!\,y!} = \frac{(n+y-1)!}{(n-1)!\,y!} = \binom{y+n-1}{n-1},$$

$$\vartheta^{-n} = \sum_{y=0}^{\infty}\binom{y+n-1}{n-1}(1-\vartheta)^{y} \Leftrightarrow \sum_{y=0}^{\infty}\binom{y+n-1}{n-1}\vartheta^{n}(1-\vartheta)^{y} = 1.$$

Σημειώστε ότι: $\sum_{x=n}^{\infty}\binom{x-1}{n-1}\vartheta^{n}(1-\vartheta)^{x-n}=1$.

$$G_y(s)=\mathbb{E}(s^y)=\sum_{y=0}^{\infty}s^y\pi(y)=\sum_{y=0}^{\infty}s^y\binom{y+n-1}{n-1}\vartheta^n(1-\vartheta)^y$$

$$=\vartheta^n\sum_{y=0}^{\infty}\binom{y+n-1}{n-1}\{s(1-\vartheta)\}^y=\vartheta^n\sum_{y=0}^{\infty}\binom{-n}{y}(-1)^y\{s(1-\vartheta)\}^y$$

$$=\vartheta^n\sum_{y=0}^{\infty}\binom{-n}{y}\{-s(1-\vartheta)\}^y=\vartheta^n\{1-s(1-\vartheta)\}^{-n}.$$

Η πιθανογεννήτρια για την τ.μ. $\tilde{y}=\sum_{i=1}^{n}y_i$ για $\tilde{n}=\sum_{i=1}^{n}n_i$ θα είναι:

$$G_{\tilde{y}}(s)=\mathbb{E}(u^{\tilde{y}})=\prod_{i=1}^{n}\mathbb{E}(u^{y_i})=\prod_{i=1}^{n}\vartheta^{n_i}\{1-u(1-\vartheta)\}^{-n_i}=\vartheta^{\tilde{n}}\{1-u(1-\vartheta)\}^{-\tilde{n}}.$$

Έτσι για την αρνητική διωνυμική έχουμε:

1. $\pi(y|\vartheta)=NB(y\mid n,\ \vartheta)=\binom{y+n-1}{n-1}\vartheta^n(1-\vartheta)^y\cdot 1(y\geq 0)$

2. $G_y(u)=\vartheta^n\left(1-u(1-\vartheta)\right)^{-n}\Rightarrow\mathbb{E}(y|\vartheta)=\Pi'_y(1)=n(1-\vartheta)\vartheta^{-1}$.

3. $y_i|\vartheta\overset{iid}{\sim}NB(\cdot|n_i,\vartheta)\Rightarrow\sum_i y_i|\vartheta\ \sim\ NB\left(\sum_i n_i,\ \vartheta\right)$.

Το μοντέλο beta – negative – binomial: Εάν $x_i|\vartheta\overset{iid}{\sim}Geo(\cdot|\vartheta), 1\leq i\leq n$,

γνωρίζουμε ότι $x|\vartheta=\sum_{i=1}^{n}x_i|\vartheta\ \sim\ NB(n,\vartheta)$ και είναι εύκολο να δούμε ότι

$\hat{\vartheta}_{MLE}=(1+\overline{x})^{-1}=\left(1+\frac{x}{n}\right)^{-1}$. Πράγματι

$$\frac{\partial}{\partial\vartheta}\log\pi(x|\vartheta)=\frac{\partial}{\partial\vartheta}\{n\log(\vartheta)+x\log(1-\vartheta)\}=\frac{n}{\vartheta}-\frac{x}{1-\vartheta}=0\Rightarrow\hat{\vartheta}=\frac{n}{n+x}$$

$$\frac{\partial^2}{\partial\vartheta^2}\log\pi(x|\vartheta)\Big|_{\vartheta=\hat{\vartheta}}=\frac{\partial}{\partial\vartheta}\left\{\frac{n}{\vartheta}-\frac{x}{1-\vartheta}\right\}\Big|_{\vartheta=\hat{\vartheta}}$$

$$= -\frac{n}{\vartheta^2} - \frac{x}{(1-\vartheta)^2}\Big|_{\vartheta=\hat{\vartheta}} = -(n+x)^2\left(\frac{1}{n}+\frac{1}{x}\right) < 0,\ x \neq 0\,.$$

Η μορφή του MLE επαληθεύεται και από το γεγονός ότι

$$\mathbb{E}[x_i|\vartheta] = \frac{1-\vartheta}{\vartheta} \Leftrightarrow \overline{x} \approx \frac{1-\vartheta}{\vartheta} \Leftrightarrow \vartheta \approx (1+\overline{x})^{-1}$$

Η EF – αναπαράσταση της $NB(n,\vartheta)$ είναι

$$\pi(x|\vartheta) = \binom{x+n-1}{n-1}\vartheta^n(1-\vartheta)^x = \binom{x+n-1}{n-1}\vartheta^n e^{x\log(1-\vartheta)}$$, δηλαδή $g(\vartheta)=\vartheta$ και

$c(\vartheta)=\log(1-\vartheta)$ με $t_{suff}(x) = x = \sum_{i=1}^{n} x_i$ που δίνει

$$\pi_{NCP}(\vartheta) \propto \vartheta^d e^{b\log(1-\vartheta)} \propto Be(\vartheta\,|\,d+1, b+1)\,.$$

Θέτοντας $\pi(\vartheta) = Be(\vartheta\,|\,a,b)$ έχουμε:

$$\pi(\vartheta\,|\,x) \propto \left\{\vartheta^{a-1}(1-\vartheta)^{b-1}\right\}\left\{\vartheta^n(1-\vartheta)^x\right\} \propto Be(\vartheta\,|\,a+n, b+x)$$

Ο σημειακός εκτιμητής ως προς τετραγωνική συνάρτηση απώλειας είναι

$$\hat{\vartheta}_{BAYES} = \mathbb{E}(\vartheta|x) = \frac{a+n}{a+b+n+x}$$

$$= \frac{a}{a+b}\left(\frac{a+b}{a+b+n+x}\right) + \frac{n}{a+b+n+x}$$

$$= \frac{a}{a+b}\left(\frac{a+b}{a+b+n+x}\right) + \frac{n}{n+x}\left(\frac{n+x}{a+b+n+x}\right)$$

$$= \mathbb{E}(\vartheta)\left(\frac{a+b}{a+b+n+x}\right) + \hat{\vartheta}_{MLE}\left(1 - \frac{a+b}{a+b+n+x}\right).$$

Παρατηρούμε ότι σε αυτή την περίπτωση το ποσοστό της κλασικής εκτίμησης στο $\hat{\vartheta}_{BAYES}$ εξαρτάται από το δείγμα μέσω της επαρκούς στατιστικής

$$\nu_n = \nu_n(x) = \frac{a+b}{a+b+n+t_{suff}(x)}\,.$$

Η prior κατανομή πρόγνωσης είναι η beta – negative – binomial κατανομή

$$BNB(x|n,a,b) \propto \binom{x+n-1}{n-1} B(a+n,b+x)\cdot 1(x\in\mathbb{N}_0).$$

Πράγματι

$$\pi(x) = \int_{\vartheta=0}^{1} \pi(x|\vartheta)\Pi(d\vartheta) = \int_{\vartheta=0}^{1} \binom{x+n-1}{n-1} \vartheta^n (1-\vartheta)^x \frac{\vartheta^{a-1}(1-\vartheta)^{b-1}}{B(a,b)} d\vartheta$$

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

Εύρεση της μέσης τιμής και διασποράς του x:

$$\mathbb{E}(x) = \mathbb{E}(\mathbb{E}(x|\vartheta)) = \mathbb{E}\left(n\frac{1-\vartheta}{\vartheta}\right) = n\mathbb{E}\left(\frac{1-\vartheta}{\vartheta}\right)$$

$$= n\int_{\vartheta=0}^{1} \left(\frac{1-\vartheta}{\vartheta}\right) \frac{\vartheta^{a-1}(1-\vartheta)^{b-1}}{B(a,b)} d\vartheta = n\frac{B(a-1,b+1)}{B(a,b)} = \frac{nb}{a-1},\ a>1.$$

$$\mathbb{V}(x) = \mathbb{V}(\mathbb{E}(x|\vartheta)) + \mathbb{E}(\mathbb{V}(x|\vartheta)) = \mathbb{V}\left(n\frac{1-\vartheta}{\vartheta}\right) + \mathbb{E}\left(n\frac{1-\vartheta}{\vartheta^2}\right)$$

$$= n^2\mathbb{V}\left(\frac{1-\vartheta}{\vartheta}\right) + n\mathbb{E}\left(\frac{1-\vartheta}{\vartheta^2}\right) = n^2\left\{\mathbb{E}\left[\left(\frac{1-\vartheta}{\vartheta}\right)^2\right] - \mathbb{E}\left[\left(\frac{1-\vartheta}{\vartheta}\right)\right]^2\right\} + n\mathbb{E}\left(\frac{1-\vartheta}{\vartheta^2}\right)$$

$$= \frac{nb}{(a-1)(a-2)}\left\{a+b+n-1+\frac{nb}{a-1}\right\},\ a>2.$$

Παρατηρείστε ότι το πραγματικό beta – ολοκλήρωμα ορίζεται σαν $B(a,b) = \int_0^1 w^{a-1}(1-w)^{b-1} dw$ για $a>0$ και $b>0$. Κατά τον υπολογισμό της $\mathbb{E}(x)$ χρησιμοποιούμε στις πράξεις μας το ολοκλήρωμα $B(a-1,b+1)$ και έτσι ζητάμε $a>1$, ενώ για τον υπολογισμό της $\mathbb{V}(x)$, έχουμε $\mathbb{E}\left(\frac{1-\vartheta}{\vartheta^2}\right) = \frac{B(a-2,b+1)}{B(a,b)}$ και έτσι θα πρέπει να έχουμε $a>2$.

Εύρεση της μέσης τιμής και διασποράς του $[y|x]$, όπου y μέλλουσα (unobserved) αρνητική διωνυμική παρατήρηση.

Πρώτα παρατηρούμε ότι $\mathbb{E}(y\,|\,x)=\mathbb{E}\left[\mathbb{E}(y\,|\,\vartheta)\,|\,x\right]$ εφόσον:

$$\mathbb{E}(y\,|\,x)=\int_Y y\pi(y\,|\,x)dy=\int_Y y\left\{\int_\Theta \pi(y\,|\,\vartheta)\pi(\vartheta\,|\,x)d\vartheta\right\}dy$$

$$=\int_\Theta\left\{\int_Y y\pi(y\,|\,\vartheta)d\ y\right\}\pi(\vartheta\,|\,x)d\vartheta=\int_\Theta \mathbb{E}(y\,|\,\vartheta)\pi(\vartheta\,|\,x)d\vartheta=\mathbb{E}\left[\mathbb{E}(y\,|\,\vartheta)\,|\,x\right]$$

Έτσι

$$\mathbb{E}(y\,|\,x)=\mathbb{E}\left(n\frac{1-\vartheta}{\vartheta}\,|\,x\right)=n\mathbb{E}\left(\frac{1-\vartheta}{\vartheta}\,|\,x\right)$$

$$=n\int_{\vartheta=0}^{1}\left(\frac{1-\vartheta}{\vartheta}\right)Be(\vartheta\,|\,a+n,b+x)d\vartheta=n\frac{B(a+n-1,b+x+1)}{B(a+n,b+x)}=\frac{n(b+x)}{a+n-1},\ a+n>1.$$

Κατά την ίδια έννοια η διασπορά γίνεται

$$\mathbb{V}(y\,|\,x)=\mathbb{V}\left(\mathbb{E}(y\,|\,\vartheta)\,|\,x\right)+\mathbb{E}\left(\mathbb{V}(y\,|\,\vartheta)\,|\,x\right)$$

$$=\frac{n(b+x)}{(a+n-1)(a+n-2)}\left\{a+b+x+2n-1+\frac{n(b+x)}{a+n-1}\right\},\ a+n>2.$$

Το μοντέλο gamma – gamma: Το συνεχές ανάλογο της γεωμετρικής κατανομής είναι η εκθετική κατανομή. Εάν $x_i\,|\,\vartheta \overset{iid}{\sim} Exp(\vartheta)=Ga(1,\vartheta), 1\le i\le n$, γνωρίζουμε ότι $x|\vartheta=\sum_{i=1}^{n}x_i\,|\,\vartheta \sim Ga(n,\vartheta)$ και εύκολα βλέπουμε ότι $\hat{\vartheta}_{MLE}=\frac{n}{\overline{x}}$. Γενικεύοντας θεωρούμε $x_i\,|\,\vartheta \overset{iid}{\sim} Ga(\,\cdot\,|\,p,\vartheta), 1\le i\le n$ για γνωστό πραγματικό $p>0$, που είναι το συνεχές ανάλογο της αρνητικής διωνυμικής κατανομής με $\hat{\vartheta}_{MLE}=\frac{p}{\overline{x}}$

Η EF – αναπαράσταση της $Ga(p,\vartheta)$ είναι

$$\pi(x_i\,|\,\vartheta)=\frac{\vartheta^p}{\Gamma(p)}x_i^{p-1}e^{-\vartheta x_i}=\left(\frac{x_i^{p-1}}{\Gamma(p)}\right)\left(\vartheta^p\right)\left(e^{-\vartheta x_i}\right),$$

δηλαδή $h(x_i)=\frac{x_i^{p-1}}{\Gamma(p)}$, $g(\vartheta)=\vartheta^p$ και $c(\vartheta)=-\vartheta$ με $t(x_i)=x_i$, με αποτέλεσμα

$$\pi_{NCP}(\vartheta)\propto(\vartheta^p)^d e^{-b\vartheta}\propto Ga(\vartheta\mid pd+1,b).$$

Η πιθανοφάνεια είναι

$$L(\vartheta\mid x_1,\ldots,x_n)=\pi(x_1,\ldots,x_n\mid\vartheta)=\prod_{i=1}^{n}\left(\frac{x_i^{p-1}}{\Gamma(p)}\right)(\vartheta^p)(e^{-\vartheta x_i})$$

$$\underset{\propto}{\vartheta}\ \vartheta^{np}e^{-n\vartheta\overline{x}}\ \propto\ Ga(\vartheta\mid np+1,n\overline{x}),\ \text{και}\ t_{suff}(x_1,\ldots,x_n)=\overline{x}.$$

Θέτοντας $\pi(\vartheta)=Ga(\vartheta\mid a,b)$ έχουμε:

$$\pi(\vartheta\mid x_1,\ldots,x_n)\propto\left\{\vartheta^{a-1}e^{-b\vartheta}\right\}\left\{\vartheta^{np}e^{-n\vartheta\overline{x}}\right\}\propto Ga(\vartheta\mid a+np,b+n\overline{x}).$$

Ο σημειακός εκτιμητής ως προς τετραγωνική συνάρτηση απώλειας είναι

$$\hat{\vartheta}_{BAYES}=\mathbb{E}(\vartheta\mid x)=\frac{a+np}{b+n\overline{x}}=\frac{a}{b}\left(\frac{b}{b+n\overline{x}}\right)+\frac{np}{b+n\overline{x}}$$

$$=\frac{a}{b}\left(\frac{b}{b+n\overline{x}}\right)+\frac{p}{\overline{x}}\left(\frac{n\overline{x}}{b+n\overline{x}}\right)=\mathbb{E}(\vartheta)\left(\frac{b}{b+n\overline{x}}\right)+\hat{\vartheta}_{MLE}\left(1-\frac{b}{b+n\overline{x}}\right).$$

Παρατηρούμε ότι και σε αυτή την περίπτωση το ποσοστό της κλασικής εκτίμησης στο $\hat{\vartheta}_{BAYES}$ εξαρτάται από το δείγμα μέσω της επαρκούς στατιστικής

$$\nu_n=\nu_n(x_1,\ldots,x_n)=\frac{b}{b+t_{suff}(x_1,\ldots,x_n)}.$$

Η prior κατανομή πρόγνωσης για μια παρατήρηση x είναι η gamma – gamma κατανομή

$$Gg(x\mid p,a,b)\propto x^{p-1}(b+x)^{-(a+p)}1(x>0)\ \text{για}\ p>0,a>0,b>0.$$

Πράγματι

$$\pi(x)=\int_{\vartheta=0}^{\infty}Ga(x\mid p,\vartheta)Ga(\vartheta\mid a,b)d\vartheta=\int_{\vartheta=0}^{\infty}\frac{\vartheta^p}{\Gamma(p)}x^{p-1}e^{-\vartheta x}\frac{b^a}{\Gamma(a)}\vartheta^{a-1}e^{-b\vartheta}d\vartheta$$

$$=\frac{b^a x^{p-1}}{\Gamma(p)\Gamma(a)}\int_{\vartheta=0}^{\infty}\vartheta^{a+p-2}e^{-(b+x)\vartheta}d\vartheta=\frac{b^a\,\Gamma(a+p)}{\Gamma(a)\Gamma(p)}\frac{x^{p-1}}{(b+x)^{a+p}}=\frac{b^a}{B(p,a)}x^{p-1}(b+x)^{-(a+p)}$$

Εύρεση της μέσης τιμής και διασποράς της x:

$$\mathbb{E}(x)=\mathbb{E}(\mathbb{E}(x\,|\,\vartheta))=\mathbb{E}\left(\frac{p}{\vartheta}\right)=p\mathbb{E}(\vartheta^{-1})=p\int_{\vartheta=0}^{\infty}\vartheta^{-1}\frac{b^a}{\Gamma(a)}\vartheta^{a-1}e^{-b\vartheta}d\vartheta$$

$$=\frac{pb^a}{\Gamma(a)}\int_{\vartheta=0}^{\infty}\vartheta^{(a-1)-1}e^{-b\vartheta}d\vartheta=\frac{pb^a}{\Gamma(a)}\frac{\Gamma(a-1)}{b^{a-1}}=\frac{pb}{a-1},\ a>1.$$

$$\mathbb{V}(x)=\mathbb{V}(\mathbb{E}(x\,|\,\vartheta))+\mathbb{E}(\mathbb{V}(x\,|\,\vartheta))=\mathbb{V}\left(\frac{p}{\vartheta}\right)+\mathbb{E}\left(\frac{p}{\vartheta^2}\right)$$

$$=p^2\mathbb{V}(\vartheta^{-1})+p\mathbb{E}(\vartheta^{-2})=p^2\left\{\mathbb{E}\left[(\vartheta^{-1})^2\right]-\mathbb{E}\left[\vartheta^{-1}\right]^2\right\}+p\mathbb{E}(\vartheta^{-2})$$

$$=p(p+1)\mathbb{E}(\vartheta^{-2})-\left\{p\mathbb{E}(\vartheta^{-1})\right\}^2=(p+1)\frac{pb^a}{\Gamma(a)}\frac{\Gamma(a-2)}{b^{a-2}}-\left\{\frac{pb}{a-1}\right\}^2$$

$$=(p+1)\frac{pb^2}{(a-1)(a-2)}-\left\{\frac{pb}{a-1}\right\}^2=\frac{pb^2}{a-1}\left(\frac{p+1}{a-2}-\frac{p}{a-1}\right),\ a>2.$$

Εύρεση της μέσης τιμής και διασποράς του $[y\,|\,x]$, όπου y μία μέλλουσα (unobserved) gamma παρατήρηση.

Μπορούμε να χρησιμοποιήσουμε ότι $\mathbb{E}(y\,|\,x)=\mathbb{E}\left[\mathbb{E}(y\,|\,\vartheta)\,|\,x\right]$ και ότι $\mathbb{V}(y\,|\,x)=\mathbb{V}(\mathbb{E}(y\,|\,\vartheta)\,|\,x)+\mathbb{E}(\mathbb{V}(y\,|\,\vartheta)\,|\,x)$. Εναλλακτικά, πολύ μπορούμε να αντικαταστήσουμε στην εξίσωση για την μέση τιμή της prior predictive $\mathbb{E}(x)=\frac{pb}{a-1}$ και τη διασπορά $\mathbb{V}(x)=\frac{pb^2}{a-1}\left(\frac{p+1}{a-2}-\frac{p}{a-1}\right)$ το posterior update των υπερπαραμέτρων $(a,b)\to(a+np,b+n\overline{x})$. Έτσι παίρνουμε

$$\mathbb{E}(y\,|\,x)=\frac{p(b+n\overline{x})}{a+np-1},$$

$$\mathbb{V}(y\,|\,x)=\frac{p(b+n\overline{x})^2}{a+np-1}\left(\frac{p+1}{a+np-2}-\frac{p}{a+np-1}\right).$$

Priors

Για να διαλέξουμε μοντέλο για prior θα πρέπει να πάρουμε υπόψη μας τα εξής

1. **Conjugate Priors**: Εάν με το συγκεκριμένο μοντέλο prior είναι δυνατόν να κάνουμε αναλυτικά υπολογισμούς για το μοντέλο μας. Σε αυτή την περίπτωση θα πρέπει να διαλέξουμε συζυγή (conjugate) prior. **Δηλαδή ανάλογα με το μοντέλο δειγματοληψίας (που θα πρέπει να ανήκει στην εκθετική οικογένεια), θα πρέπει να διαλέξουμε prior τέτοιο ώστε a-priori και a-posteriori κατανομές να είναι μέλη της ίδιας οικογένειας κατανομών.** Αν ο prior είναι μη συζυγής το πιθανότερο είναι ότι ακόμα και αν καταφέρουμε να κάνουμε αναλυτικούς υπολογισμούς θα καταλήξουμε σε κάποια μη αναγνωρίσιμη (nonstandard) posterior και μάλιστα τις περισσότερες φορές, δεν θα γνωρίζουμε ούτε καν την σταθερά κανονικοποίησης. Τότε η εύρεση οποιασδήποτε σημειακής εκτίμησης θα πρέπει να γίνει μέσο προσομοίωσης Markov Chain Monte Carlo (MCMC).

2. **Informative Priors**: Εάν με το συγκεκριμένο μοντέλο για prior θέλουμε να εκφράσουμε πεποιθήσεις που βασίζονται είτε σε προηγούμενα πειράματα είτε σε γνώμες ειδικών (elicitation). Οι a-priori κατανομές που προκύπτουν με αυτό τον τρόπο λέγονται informative και μπορεί να είναι είτε συζυγείς είτε μη συζυγείς.

3. **Noninformative Priors**: Εάν με το συγκεκριμένο μοντέλο για prior θέλουμε να εκφράσουμε την απουσία είτε την αμελητέα επιρροή της a-priori πληροφορίας, είτε απλά θέλουμε να μην εισάγουμε στο μοντέλο μας αρχική πληροφορία για την άγνωστη παράμετρο Τέτοιες a-priori κατανομές ονομάζονται αντικειμενικές (objective), reference, flat ή vague. Όπως θα δούμε στην συνέχεια τέτοιοι priors μπορεί να μην είναι καν κατανομές αλλά απλώς όρια του πυρήνα κατανομών. Επίσης με τη σειρά τους μπορεί να είναι συζυγής ή μη συζυγής. Στην μονοδιάστατη περίπτωση, $\Theta \subseteq \mathbb{R}$, τα Jeffrey's priors χρησιμοποιούνται ευρέως.

Nonconjugate priors

Έστω Poisson μοντέλο δειγματοληψίας $x_i|\vartheta \overset{iid}{\sim} Po(\cdot|\vartheta),\ 1\le i\le n$. Σε αυτή τη περίπτωση, γνωρίζουμε ότι ο αντίστοιχος συζυγής prior είναι $\pi(\vartheta)=Ga(\vartheta|p,q)$ και επειδή η πιθανοφάνεια είναι της μορφής $\pi(x|\vartheta)\propto e^{-n\vartheta}\vartheta^{n\bar{x}}$, παίρνουμε,

$\pi(\vartheta|x)=Ga(\vartheta|p+n\bar{x},q+n)$.

Εάν όμως έχουμε την αρχική πληροφορία ότι $0 < \vartheta < 1$, θα ήταν λογικό να θέσουμε $\pi(\vartheta) = Be(\vartheta | p, q)$. Η posterior τότε παίρνει τη μορφή

$$\pi(\vartheta | x) \propto e^{-n\vartheta} \vartheta^{p+n\bar{x}-1} (1-\vartheta)^{q-1},$$

που για $q \neq 1$ είναι ο πυρήνας μίας μη αναγνωρίσιμης κατανομής. Δηλαδή $\pi(\vartheta | x) = C \cdot e^{-n\vartheta} \vartheta^{p+n\bar{x}-1} (1-\vartheta)^{q-1}$, με σταθερά κανονικοποίησης $C = \left\{ \int_0^1 e^{-n\vartheta} \vartheta^{p+n\bar{x}-1} (1-\vartheta)^{q-1} \, d\vartheta \right\}^{-1}$ άγνωστη[1] και αυτή.

Στην ίδια κατάσταση (δηλαδή με μη αναγνωρίσιμη posterior) θα βρεθούμε εάν θέσουμε σαν prior $\pi(\vartheta) = LN\left(\vartheta | m, c^{-1}\right)$ (lognormal) για $\vartheta > 0$. Τότε

$\pi(\vartheta) = LN\left(\vartheta | m, c^{-1}\right) = \sqrt{\dfrac{c}{2\pi}} \dfrac{1}{\vartheta} \exp\left\{ -\dfrac{c}{2} \left(\log(\vartheta) - m \right)^2 \right\}$ που δίνει

$$\pi(\vartheta | x) \; \propto \; \vartheta^{n\bar{x}-1} \exp\left\{ -\frac{c}{2} \left(\ln(\vartheta) - m \right)^2 - n\vartheta \right\}.$$

The Exponential Family of distributions

Θεωρούμε ότι το μοντέλο δειγματοληψίας ανήκει στην εκθετική οικογένεια κατανομών (exponential family$-EF$) όταν:

$$\pi(x_i | \vartheta) \propto h(x_i) \exp\left[c(\vartheta) t(x_i) \right],$$

και ο χώρος καταστάσεων της τ.μ. $[x_i | \vartheta]$ δεν εξαρτάται από την παράμετρο ϑ. Η σταθερά κανονικοποίησης είναι $g(\vartheta) = \left\{ \int_{\mathbb{R}} h(u) \exp\left[c(\vartheta) t(u) \right] du \right\}^{-1}$ και

$$\pi(x_i | \vartheta) = \frac{h(x_i) \exp\left[c(\vartheta) t(x_i) \right]}{\int_{\mathbb{R}} h(u) \exp\left[c(\vartheta) t(u) \right] du} = h(x_i) g(\vartheta) \exp\left[c(\vartheta) t(x_i) \right].$$

Ονομάζουμε την συνάρτηση $c(\vartheta)$ φυσική παράμετρο της $\pi(\cdot | \vartheta) \in EF$.

---

[1] Στην ειδική περίπτωση $q = 1$, έχουμε $\pi(\vartheta | x) \propto e^{-n\vartheta} \vartheta^{p+n\bar{x}-1} \propto Ga(\vartheta | p + n\bar{x}, n)$.

Θεωρούμε ότι η οικογένεια των ομοιόμορφων κατανομών στο διάστημα $(0,\vartheta)$ για $\vartheta>0$ δεν ανήκει στην εκθετική οικογένεια, εφόσον ο χώρος των καταστάσεων εξαρτάται από το ϑ.

Στη βιβλιογραφία, πολλές φορές η $\pi(\,\cdot|\vartheta)\in EF$ αναπαριστάται ισοδύναμα σαν:

$$\pi(\,\cdot|\vartheta)\in EF \Leftrightarrow \pi(x|\vartheta)=h(x)\exp\left[c(\vartheta)t(x)-A(\vartheta)\right],$$

με $A(\vartheta)=\log\int_{\mathbb{R}} h(u)\exp\left[c(\vartheta)t(u)\right]du$.

Η πιθανοφάνεια μιας δειγματοληπτικής κατανομής που ανήκει στην EF, για $x=(x_1,\ldots,x_n)$, δίνεται από την σχέση:

$$\pi(x|\vartheta)=h(x)g(\vartheta)^n\exp\left[c(\vartheta)t(x)\right] \overset{\vartheta}{\propto} g(\vartheta)^n\exp\left[c(\vartheta)t(x)\right],$$

όπου $t(x)=\sum_{i=1}^n t(x_i)$ και $h(x)=\prod_{i=1}^n h(x_i)$.

Είναι τότε προφανές από την παραγοντοποίηση της πιθανοφάνειας κατά Fisher–Neyman $\pi(x|\vartheta)=h_{FN}(x)g_{FN}(t(x),\vartheta)$ ότι $h_{FN}(x)=h(x)$ και $g_{FN}(t(x),\vartheta)=g(\vartheta)^n\exp\left[c(\vartheta)t(x)\right]$. Ή ότι η δειγματική συνάρτηση $t(x)=t_{suff}(x)=\sum_{i=1}^n t(x_i)$ είναι επαρκής στατιστική.

Ο φυσικός συζυγής prior (Natural Conjugate Prior $-NCP$)

Θέτουμε σαν NCP prior μοντέλο την συναρτησιακή έκφραση του $\vartheta-$ πυρήνα της πιθανοφάνειας

$$\pi(x|\vartheta)=h(x)g(\vartheta)^n\exp\left[c(\vartheta)t(x)\right] \overset{\vartheta}{\propto} g(\vartheta)^n\exp\left[c(\vartheta)t(x)\right]$$

$$\Rightarrow \pi_{NCP}(\vartheta)=\pi^*(\vartheta\,|\,d,b) \overset{\vartheta}{\propto} g(\vartheta)^d\exp\left[c(\vartheta)b\right],$$

όπου b και d υπερπαράμετροι.

Αντίστροφα η EF πιθανοφάνεια μπορεί να αναπαρασταθεί χρησιμοποιώντας NCP

$$\pi\left(x|\vartheta\right) \overset{\vartheta}{\propto} g\left(\vartheta\right)^{n}\exp\left[c\left(\vartheta\right)t\left(x\right)\right] \overset{\vartheta}{\propto} \pi^{*}\left(\vartheta\,|\,n,t\left(x\right)\right).$$

Η NCP posterior τότε έχει την αναπαράσταση

$$\pi_{NCP}\left(\vartheta|x\right) \propto \left\{g\left(\vartheta\right)^{d}\exp\left[bc\left(\vartheta\right)\right]\right\} \times \left\{g\left(\vartheta\right)^{n}\exp\left[c\left(\vartheta\right)t\left(x\right)\right]\right\}$$

$$= g\left(\vartheta\right)^{d+n}\exp\left[\left(b+t\left(x\right)\right)c\left(\vartheta\right)\right] \propto \pi^{*}\left(\vartheta\,|\,d+n,b+t\left(x\right)\right).$$

Δηλαδή για την οικογένεια prior κατανομών $\pi_{NCP}\left(\vartheta\right)=\pi\left(\vartheta\,|\,d,b\right)$ έχουμε posterior update $\pi_{NCP}\left(\vartheta|x\right)=\pi\left(\vartheta\,|\,d+n,b+t\left(x\right)\right)$ έτσι:

$$\pi_{NCP}\left(\vartheta\right)=\frac{g\left(\vartheta\right)^{d}\exp\left[b\cdot c\left(\vartheta\right)\right]}{\int_{\Theta} g\left(u\right)^{d}\exp\left[b\cdot c\left(u\right)\right]du},$$

$$\pi_{NCP}\left(\vartheta|x\right)=\frac{g\left(\vartheta\right)^{d+n}\exp\left[\left(b+t\left(x\right)\right)c\left(\vartheta\right)\right]}{\int_{\Theta} g\left(u\right)^{d+n}\exp\left[\left(b+t\left(x\right)\right)c\left(u\right)\right]du}.$$

Προφανώς η posterior ανήκει στην ίδια οικογένεια όπως και η prior. Η επιλογή δειγματοληπτικής κατανομή από την EF δεν μας περιορίζει δραστικά, εφόσον κατανομές όπως η normal, exponential, gamma, inverse—gamma, binomial, geometric negative binomial, poisson, beta, κλπ ανήκουν στην εκθετική οικογένεια[2].

NCP για binomial data: Έστω ότι $y_i\,|\,\vartheta \overset{iid}{\sim} Bin\left(\cdot|1,\vartheta\right)$, $1\le i\le n$ τότε $x|\vartheta=\sum_{i=1}^{n} y_i\,|\,\vartheta \sim Bin\left(\cdot|n,\vartheta\right)$. Η εκθετική αναπαράσταση της Binomial είναι

$$Bin\left(x|n,\vartheta\right)=\binom{n}{x}\vartheta^{x}\left(1-\vartheta\right)^{n-x}=\binom{n}{x}\left(1-\vartheta\right)^{n}\exp\left[x\log\left(\frac{\vartheta}{1-\vartheta}\right)\right]$$

με $h\left(x\right)=\binom{n}{x},\quad g\left(\vartheta\right)=\left(1-\vartheta\right)^{n},\quad t\left(x\right)=x$. και φυσική παράμετρο $c\left(\vartheta\right)=\log\left(\frac{\vartheta}{1-\vartheta}\right)$.

[2] Η student $-t_{\nu}$ για $\nu<\infty$ και η ομοιόμορφη με άγνωστες παραμέτρους δεν ανήκουν στην εκθετική οικογένεια.

Το NCP μοντέλο είναι: $\pi_{NCP}(\vartheta) \propto g(\vartheta)^d \exp\left[bc(\vartheta)\right]$

$$\pi_{NCP}(\vartheta) \propto (1-\vartheta)^{nd} \exp\left[b\log\left(\frac{\vartheta}{1-\vartheta}\right)\right] = \vartheta^b (1-\vartheta)^{nd-b} \propto Be(\vartheta|b+1, nd-b+1)$$

με posterior

$$\pi_{NCP}(\vartheta|x) \propto \left\{\vartheta^b (1-\vartheta)^{nd-b}\right\} \times \left\{\vartheta^x (1-\vartheta)^{n-x}\right\} \propto \vartheta^{b+x} (1-\vartheta)^{nd-b+n-x}$$

$$\propto Be(\vartheta|b+x+1, nd-b+n-x+1)$$

Θέτοντας $\begin{Bmatrix} p = b+1 \\ q = nd-b+1 \end{Bmatrix}$ παίρνουμε $\begin{Bmatrix} b+x+1 = x+p \\ nd-b+n-x+1 = q+n-x \end{Bmatrix}$, και τελικά καταλήγουμε στις γνωστές σχέσεις:

$$\begin{array}{ll} prior & \pi(\vartheta) = Be(\vartheta|p,q) \\ likelihood & \pi(x|\vartheta) = Bin(x|n,\vartheta) \\ posterior & \pi(\vartheta|x) = Be(\vartheta|p+x, q+n-x) \end{array}$$

<u>Άσκηση</u>: Επαναλάβετε την προηγούμενη ανάλυση χρησιμοποιώντας Bernoulli παρατηρήσεις

$$\pi(y_i|\vartheta) = \vartheta^{y_i}(1-\vartheta)^{1-y_i} = (1-\vartheta)\exp\left\{y_i \log\left(\frac{\vartheta}{1-\vartheta}\right)\right\}$$

$$\pi(y_1,\ldots,y_n|\vartheta) = (1-\vartheta)^n \exp\left\{\log\left(\frac{\vartheta}{1-\vartheta}\right)\sum_{i=1}^n y_i\right\}$$

$$h(y)=1,\ g(\vartheta)=(1-\vartheta),\ t(y_1,\ldots,y_n)=\sum_{i=1}^n y_i,\ c(\vartheta)=\log\left(\frac{\vartheta}{1-\vartheta}\right)$$

etc…

Εάν επιλέξουμε $\pi(\vartheta) \propto \vartheta^{-1}(1-\vartheta)^{-1}$ (που είναι improper prior στο $(0,1)$), noninformative και συζυγής, θα έχουμε

$$\pi(\vartheta|x) \propto \vartheta^{-1}(1-\vartheta)^{-1} \times \left\{\vartheta^{n\bar{x}}(1-\vartheta)^{n-n\bar{x}}\right\} \propto Be(\vartheta|n\bar{x}, n-n\bar{x})$$

από όπου και $\mathbb{E}(\vartheta|x) = \hat{\vartheta}_{MLE}$.

NCP για poisson data: Έστω ότι $x_i|\vartheta \overset{iid}{\sim} Po(\cdot|\vartheta)$, $1 \le i \le n$

$$\pi(x_i|\vartheta) = Po(x_i|\vartheta) = \frac{1}{x_i!} e^{-\vartheta} e^{x_i \log(\vartheta)}$$

από όπου $h(x_i) = \frac{1}{x_i!}$, $g(\vartheta) = e^{-\vartheta}$, $c(\vartheta) = \log(\vartheta)$, $t(x_i) = x_i$

$$\pi(x|\vartheta) = \left(\prod_{i=1}^{n} x_i!\right)^{-1} \left(e^{-\vartheta}\right)^n e^{\log(\vartheta)\sum_{i=1}^n x_i} \propto e^{-n\vartheta} e^{\log(\vartheta) t(x)}$$ με $t(x) = \sum_{i=1}^n x_i$.

Το NCP μοντέλο θα είναι

$$\pi_{NCP}(\vartheta) \propto g(\vartheta)^d \exp\left[b\, c(\vartheta)\right] = \left(e^{-\vartheta}\right)^d e^{b\log(\vartheta)} = \vartheta^b e^{-d\vartheta} \propto Ga(\vartheta|b+1, d)$$

με posterior

$$\pi_{NCP}(\vartheta|x) \propto \vartheta^b e^{-d\vartheta} \times e^{-n\vartheta} \vartheta^{t(x)} \propto \vartheta^{b+t(x)} e^{-(n+d)\vartheta} \propto Ga\left(\vartheta|b+t(x)+1, n+d\right)$$

Θέτοντας $\begin{Bmatrix} p = b+1 \\ q = d \end{Bmatrix} \Rightarrow \begin{Bmatrix} b+t(x)+1 = p+t(x) \\ n+d = q+n \end{Bmatrix}$

Τελικά επαληθεύουμε ότι:

$$\begin{array}{ll} prior & \pi(\vartheta) = Ga(\vartheta|p,q) \\ likelihood & \pi(x|\vartheta) = \prod_{i=1}^{n} Po(x_i|\vartheta) \\ posterior & \pi(\vartheta|x) = Ga\left(\vartheta|p+t(x), q+n\right) \end{array}$$

NCP για gamma data: Έχουμε δύο περιπτώσεις

1. Γνωστή shape παράμετρος και άγνωστο rate $x_i|\vartheta \overset{iid}{\sim} Ga(\cdot|p,\vartheta)$, $1 \le i \le n$. Τότε

$$\pi(x_i|\vartheta) = Ga(x_i|p,\vartheta) = \frac{\vartheta^p}{\Gamma(p)} x_i^{p-1} e^{-\vartheta x_i}$$

από όπου $h(x_i) = \frac{x_i^{p-1}}{\Gamma(p)}$, $g(\vartheta) = \vartheta^p$, $c(\vartheta) = -\vartheta$, $t(x_i) = x_i$.

$$\pi_{NCP}\left(\vartheta\right) \propto g\left(\vartheta\right)^{d} \exp\left[b\, c\left(\vartheta\right)\right] = \vartheta^{p\,d} e^{-b\vartheta} \propto Ga\left(\vartheta \mid p\,d+1, b\right)$$

$$\pi\left(x|\vartheta\right) = \prod_{i=1}^{n} \frac{\vartheta^{p}}{\Gamma\left(p\right)} x_i^{p-1} e^{-\vartheta x_i} = \frac{\vartheta^{n\,p}}{\Gamma\left(p\right)^{n}} \left\{\prod_{i=1}^{n} x_i\right\}^{p-1} e^{-\vartheta n \bar{x}} \overset{\vartheta}{\propto} \vartheta^{n\,p} e^{-\vartheta n \bar{x}}$$ με $t\left(x\right) = n\,\bar{x}$.

$$\pi_{NCP}\left(\vartheta|x\right) \propto \vartheta^{p\,d} e^{-b\vartheta} \times \vartheta^{n\,p} e^{-\vartheta t\left(x\right)} \propto \vartheta^{p\left(d+n\right)} e^{-\left(b+t\left(x\right)\right)\vartheta} \propto Ga\left(\vartheta \mid p\left(d+n\right)+1, b+t\left(x\right)\right).$$

2. Άγνωστη shape παράμετρο και γνωστό rate $x_i \mid \vartheta \overset{iid}{\sim} Ga\left(\cdot \mid \vartheta, q\right)$, $1 \le i \le n$. Τότε

$$\pi\left(x_i \mid \vartheta\right) = Ga\left(x_i \mid \vartheta, q\right) = \frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)} x_i^{\vartheta-1} e^{-q\,x_i} = \left(\frac{e^{-q\,x_i}}{x_i}\right)\left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right) \exp\left(\vartheta \log\left(x_i\right)\right),$$

από όπου $h\left(x_i\right) = \dfrac{e^{-q\,x_i}}{x_i},\ g\left(\vartheta\right) = \dfrac{q^{\vartheta}}{\Gamma\left(\vartheta\right)},\ c\left(\vartheta\right) = \vartheta,\ t\left(x_i\right) = \log\left(x_i\right)$.

$$\pi_{NCP}\left(\vartheta\right) \propto g\left(\vartheta\right)^{d} \exp\left(b\, c\left(\vartheta\right)\right) = \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{d} \exp\left(b\,\vartheta\right) \propto \pi\left(\vartheta \mid d, b\right)$$

$$\pi\left(x|\vartheta\right) = \prod_{i=1}^{n} \frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)} x_i^{\vartheta-1} e^{-q\,x_i} = \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{n} \left\{\prod_{i=1}^{n} x_i\right\}^{\vartheta-1} e^{-q\,n\,\bar{x}} \overset{\vartheta}{\propto} \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{n} \left\{\prod_{i=1}^{n} x_i\right\}^{\vartheta}$$

$$\overset{\vartheta}{\propto} \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{n} \exp\left(\vartheta t_{suff}\left(x\right)\right) \propto \pi\left(\vartheta \mid n, t_{suff}\left(x\right)\right)$$, με $t_{suff}\left(x\right) = \sum_{i=1}^{n} \log\left(x_i\right)$

$$\pi_{NCP}\left(\vartheta|x\right) \propto \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{d} \exp\left(b\,\vartheta\right) \times \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{n} \exp\left(\vartheta\, t_{suff}\left(x\right)\right)$$

$$\propto \left(\frac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{d+n} \exp\left(\vartheta\left(b + t_{suff}\left(x\right)\right)\right) \propto \pi\left(\vartheta | d+n, b+t_{suff}\left(x\right)\right).$$

Όταν λοιπόν έχουμε άγνωστη shape παράμετρο και γνωστό rate, το prior μοντέλο $\pi_{NCP}\left(\vartheta\right) \propto \left(\dfrac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{d} \exp\left(b\,\vartheta\right)$ και η αντίστοιχη posterior $\pi_{NCP}\left(\vartheta|x\right) \propto \left(\dfrac{q^{\vartheta}}{\Gamma\left(\vartheta\right)}\right)^{d+n} \exp\left(\vartheta\left(b + t_{suff}\left(x\right)\right)\right)$ ανήκουν σε **nonstandard οικογένεια κατανομών.**

Άσκηση: Δείξτε ότι η πυκνότητα $\pi(\vartheta \mid d,b) \propto \left(\frac{q^{\vartheta}}{\Gamma(\vartheta)}\right)^{d} \exp(b\,\vartheta)$, για $d>0$ είναι **log-concave**, δηλαδή ότι $\frac{d^2}{d\vartheta^2}\log\pi(\vartheta \mid d,b)<0$.

$$\frac{d^2}{d\vartheta^2}\log\pi(\vartheta \mid d,b) = -d\,\frac{d^2}{d\vartheta^2}\log\Gamma(\vartheta) = -d\,\Psi'(\vartheta).$$

Γνωρίζουμε ότι $\Psi'(\vartheta) \triangleq \frac{d^2}{d\vartheta^2}\log\Gamma(\vartheta) = \sum_{k=0}^{\infty}\frac{1}{(\vartheta+k)^2},\ \forall\,\vartheta\in\mathbb{R}\setminus\mathbb{Z}^{\le 0}$ και έτσι για $d>0$

$$\frac{d^2}{d\vartheta^2}\log\pi(\vartheta \mid d,b)<0 \Leftrightarrow \log\pi(\vartheta \mid d,b) = f_{\cap}(\vartheta).$$

Άσκηση: Δίνεται το παραμετρικό μοντέλο $x_i \mid \vartheta \overset{iid}{\sim} Be(\,\cdot \mid \vartheta, q),\ 1\le i\le n$, για γνωστή shape2 παράμετρο $q>0$. Να δειχθεί ότι η $\pi_{NCP}(\vartheta)$ είναι **log-concave** πυκνότητα.

$$\pi(x_i \mid \vartheta) = Be(x_i \mid \vartheta, q) = \frac{\Gamma(\vartheta+q)}{\Gamma(\vartheta)\Gamma(q)}\,x_i^{\vartheta-1}(1-x_i)^{q-1}\,1(0<x_i<1)$$

$$=\left(\frac{(1-x_i)^{q-1}}{x_i\,\Gamma(q)}\right)\left(\frac{\Gamma(\vartheta+q)}{\Gamma(\vartheta)}\right)\exp\left(\vartheta\log(x_i)\right) \Rightarrow \pi(\,\cdot \mid \vartheta)\in EF\ ,$$

με $h(x_i) = \frac{(1-x_i)^{q-1}}{x_i\,\Gamma(q)},\ g(\vartheta) = \frac{\Gamma(\vartheta+q)}{\Gamma(\vartheta)},\ c(\vartheta)=\vartheta,\ t(x_i)=\log(x_i)$.

Τότε $\pi_{NCP}(\vartheta) \propto \left(\frac{\Gamma(\vartheta+q)}{\Gamma(\vartheta)}\right)^{d}\exp(b\,\vartheta) \propto \pi(\vartheta \mid d,b)$ ενώ

$\pi_{NCP}(\vartheta \mid x) = \pi\left(\vartheta \mid d+n,\ b+t_{suff}(x)\right)$ με $t_{suff}(x) = \sum_{i=1}^{n}\log(x_i)$.

Για την πυκνότητα $\pi(\vartheta \mid d,b) \propto \left(\frac{\Gamma(\vartheta+q)}{\Gamma(\vartheta)}\right)^{d}\exp(b\,\vartheta),\ \vartheta>0,\ \forall q>0$ έχουμε

$\frac{d^2}{d\vartheta^2}\log\pi(\vartheta \mid d,b) = d\left\{\Psi'(\vartheta+q)-\Psi'(\vartheta)\right\}<0$ για $d>0$.

Πράγματι $\Psi'(\vartheta) \triangleq \frac{d^2}{d\vartheta^2}\log\Gamma(\vartheta) = \sum_{k=0}^{\infty}\frac{1}{(\vartheta+k)^2},\ \forall\,\vartheta\in\mathbb{R}\setminus\mathbb{Z}^{\le 0} \Rightarrow$

$\Psi''(\vartheta) = -2\sum_{k=0}^{\infty}\frac{1}{(\vartheta+k)^3} < 0$ για κάθε $\vartheta > 0 \Rightarrow \Psi'(\vartheta)\downarrow$ για κάθε $\vartheta > 0$.

<u>Άσκηση</u>: Δείξτε ότι η πυκνότητα $\pi(\vartheta \mid \rho,\varphi,m) \propto \vartheta^{\varphi}\frac{\rho^{\vartheta}}{\Gamma(\vartheta)^m}$, $\vartheta > 0, \varphi > 0, m > 0$, είναι **log—concave.**

Προφανώς $\frac{d^2}{d\vartheta^2}\log\pi(\vartheta \mid \rho,\varphi,m) = -\left\{\frac{\varphi}{\vartheta^2} + m\Psi'(\vartheta)\right\} < 0$.

Η κατανομή Pareto ορίζεται σαν $Pa(\vartheta \mid a,b) \propto \vartheta^{-(a+1)}1(\vartheta > b),\ a,b > 0$, με σταθερά κανονικοποίησης $C^{-1} = \int_{\mathbb{R}} \vartheta^{-(a+1)}1(\vartheta > b)\,d\vartheta = \int_{\vartheta=b}^{\infty} \vartheta^{-(a+1)}d\vartheta = \left(ab^a\right)^{-1}$, από όπου $Pa(\vartheta \mid a,b) = ab^a\vartheta^{-(a+1)}1(\vartheta > b)$. Η Pareto όταν η shape παράμετρος b είναι άγνωστη, δεν ανήκει στην εκθετική οικογένεια (ο χώρος καταστάσεων εξαρτάται από τις άγνωστες παραμέτρους).

Εύκολα μπορούμε να διαπιστώσουμε ότι η κατανομή pareto δίνεται από τη μίξη $Pa(\vartheta \mid a,b) = \int_{u=0}^{\infty} Exp(\vartheta - b \mid u)Ga(u \mid a,b)\,du$.

Πράγματι

$$\int_{u=0}^{\infty} Exp(\vartheta - b \mid u)Ga(u \mid a,b)\,du = \int_{u=0}^{\infty} ue^{-u(\vartheta-b)}1(\vartheta - b > 0)\frac{b^a}{\Gamma(a)}u^{a-1}e^{-bu}\,du$$

$= \frac{b^a}{\Gamma(a)}1(\vartheta - b > 0)\int_{u=0}^{\infty} u^a e^{-u\vartheta}\,du$, θέτουμε $u\vartheta = v$

$$= \frac{b^a}{\Gamma(a)}1(\vartheta > b)\int_{u=0}^{\infty}\left(\frac{v}{\vartheta}\right)^a e^{-v}\frac{du}{\vartheta} = \frac{b^a}{\Gamma(a)}\frac{\Gamma(a+1)}{\vartheta^{a+1}}1(\vartheta > b) = \frac{ab^a}{\vartheta^{a+1}}1(\vartheta > b).$$

Η $n-$οστή ροπή της pareto υπάρχει μόνο για $a > n$

$$\mathbb{E}(\vartheta^n) = \int_{\vartheta=-\infty}^{\infty} \vartheta^n ab^a\vartheta^{-(a+1)}1(\vartheta > b)\,d\vartheta = ab^a\int_{\vartheta=b}^{\infty}\vartheta^{n-(a+1)}d\vartheta = \frac{ab^n}{a-n},$$

έτσι η $\mathbb{E}(\vartheta)=\dfrac{ab}{a-1}$ υπάρχει μόνο για $a>1$ και αντίστοιχα η διασπορά $\mathbb{V}(\vartheta)$ υπάρχει μόνο όταν $a>2$ με $\mathbb{V}(\vartheta)=\dfrac{ab^2}{(a-1)(a-2)}$.

<u>Άσκηση</u> (**Συζυγής ανάλυση εκτός EF**) Η ομοιόμορφη κατανομή όταν τουλάχιστον μία από τις παραμέτρους της είναι άγνωστη δεν ανήκει στην εκθετική οικογένεια (ο χώρος καταστάσεων εξαρτάται από τις άγνωστες παραμέτρους). Δείξτε ότι συζυγής prior για το μοντέλο $x_i|\vartheta \overset{iid}{\sim} \mathcal{U}(\cdot|0,\vartheta),\ i=1,\ldots,n$ είναι η κατανομή Pareto, δηλαδή $\pi(\vartheta)\propto \vartheta^{-(a+1)}1(\vartheta>b)$, και εκτιμήστε το ϑ κάτω από τετραγωνική συνάρτηση απώλειας.

Γνωρίζουμε ότι στην περίπτωση $x_i|\vartheta \overset{iid}{\sim} \mathcal{U}(\cdot|0,\vartheta),\ i=1,\ldots,n$ έχουμε

$$\pi(x|\vartheta) = \prod_{i=1}^{n}\mathcal{U}(x_i|0,\vartheta)=\prod_{i=1}^{n}\vartheta^{-1}1(0\le x_i\le\vartheta) \overset{\vartheta}{\propto} \vartheta^{-n}1\left(\vartheta\ge x_{(n)}\right)$$

Εμφανώς $\pi(x|\vartheta) \overset{\vartheta}{\propto} \vartheta^{-n}1\left(\vartheta\ge x_{(n)}\right) \overset{\vartheta}{\propto} Pa\left(\vartheta|n-1,x_{(n)}\right)$

$h_{FN}(x)=1,\ g_{FN}(t(x),\vartheta)=\vartheta^{-n}1\left(\vartheta\ge x_{(n)}\right) \Rightarrow t_{suff}(x)=x_{(n)}$ [3] και $\hat{\vartheta}_{MLE}=x_{(n)}$.

Τότε θέτοντας $\pi(\vartheta)=Pa(\vartheta|a,b)\propto\vartheta^{-(a+1)}1(\vartheta>b)$ έχουμε

$$\pi(\vartheta|x)\propto\vartheta^{-(a+1)}1(\vartheta>b)\times\vartheta^{-n}1\left(\vartheta\ge x_{(n)}\right)=\vartheta^{-(a+n+1)}1\left(\vartheta>\max\left\{b,x_{(n)}\right\}\right)$$

$$\propto Pa\left(\vartheta|a+n,\max\left\{b,x_{(n)}\right\}\right),$$

που δίνει

$$E(\vartheta|x)=\frac{(a+n)\max\left\{b,x_{(n)}\right\}}{a+n-1}.$$

Ο εκτιμητής κατά Bayes σε αυτή την **περίπτωση δεν μπορεί να αναπαρασταθεί** σαν κυρτός γραμμικός συνδυασμός μεταξύ των prior πεποιθήσεων και του εκτιμητή μεγίστης πιθανοφάνειας

[3] The max order statistic

$$\mathbb{E}(\vartheta|x)=\left\{\mathbb{E}(\vartheta)v_n+\left(\frac{b}{x_{(n)}}\right)x_{(n)}(1-v_n)\right\}1\left(b>x_{(n)}\right)+\left\{\left(\frac{x_{(n)}}{b}\right)\mathbb{E}(\vartheta)v_n+x_{(n)}(1-v_n)\right\}1\left(b\le x_{(n)}\right)$$

Αριθμητικό παράδειγμα: Υποθέτουμε ότι $[x_i|\vartheta]\overset{iid}{\sim}\mathcal{U}(\cdot|0,\vartheta),\ i=1,\ldots,n$ και από προηγούμενα πειράματα κρίνουμε ότι $\pi(\vartheta)=Pa(\vartheta|2.5,4.0)$. Παρατηρούμε δείγμα μεγέθους $n=10$

y = $(0.4,1.0,1.5,1.7,2.0,2.1,3.1,3.7,4.3,4.9)$, από όπου $y_{(10)}=4.3$

Η πιθανοφάνεια είναι $\pi(y|\vartheta)\overset{\vartheta}{\propto}\vartheta^{-10}1(\vartheta\ge 4.3)\overset{\vartheta}{\propto}Pa(\vartheta|9.0,4.3)$

Η posterior είναι $\pi(\vartheta|y)=Pa(\vartheta|12.5,4.3)$

$\mathbb{E}(\vartheta)=6.667$ και $\mathbb{E}(\vartheta|y)=5.326$

$\sqrt{\mathbb{V}(\vartheta)}=5.963$ και $\sqrt{\mathbb{V}(\vartheta|y)}=0.465$.

<u>Άσκηση</u>

Δείξτε ότι συζυγής prior για το μοντέλο $[x_i|\vartheta]\overset{iid}{\sim}Pa(\cdot|\vartheta,\beta),\ i=1,\ldots,n$ είναι η κατανομή gamma (επειδή το β είναι γνωστό θα έχουμε ότι $Pa(\cdot|\vartheta,\beta)\in EF$) και εκτιμήστε το ϑ κάτω από τετραγωνική συνάρτηση απώλειας.

$$\pi(x_i|\vartheta)=Pa(x_i|\vartheta,\beta)=\vartheta\beta^{\vartheta}x_i^{-(\vartheta+1)}1(x_i>\beta)=\left\{x_i^{-1}1(x_i>\beta)\right\}\left\{\vartheta\beta^{\vartheta}\right\}e^{-\vartheta\log(x_i)}$$

$$\pi(x|\vartheta)=\vartheta^n\beta^{n\vartheta}\xi^{-(\vartheta+1)}1\left(x_{(1)}>\beta\right),\ \xi=\prod_{i=1}^n x_i$$

Εμφανώς

$$\pi(x|\vartheta)\overset{\vartheta}{\propto}\vartheta^n\beta^{n\vartheta}\xi^{-\vartheta}=\vartheta^n\exp\{-\vartheta\psi\}\overset{\vartheta}{\propto}Ga(\vartheta|n+1,\psi),\ \psi=\log\xi-n\log\beta,$$

$$\pi_{NCP}(\vartheta)\overset{\vartheta}{\propto}\left\{\vartheta\beta^{\vartheta}\right\}^d e^{-\vartheta b\log(x_i)}=\vartheta^d e^{\vartheta d\log(\beta)}e^{-\vartheta b\log(x_i)}=\vartheta^d\exp\left\{-\vartheta\left(b\log(x_i)-d\log(\beta)\right)\right\}$$

$\overset{\vartheta}{\propto}\mathcal{G}\left(\vartheta|d+1,\,b\log(x_i)-d\log(\beta)\right)$ με $b\log(x_i)-d\log(\beta)>0$

και θέτοντας $\pi(\vartheta)=Ga(\vartheta|p,q)\propto\vartheta^{p-1}e^{-q\vartheta}$, παίρνουμε

$\pi\left(\vartheta|x\right)\propto\vartheta^{p-1}e^{-q\vartheta}\times\vartheta^{n}e^{-\psi\vartheta}=\vartheta^{p+n-1}e^{-(q+\psi)\vartheta}\propto Ga\left(\vartheta\,|\,p+n,q+\psi\right)$.

$$\mathbb{E}\left(\vartheta|x\right)=\frac{p+n}{q+\psi}=\left(\frac{p}{q}\right)\left(\frac{q}{q+\psi}\right)+\left(\frac{n}{\psi}\right)\left(1-\frac{q}{q+\psi}\right).$$

<u>Άσκηση (**Μη συζυγής περίπτωση)**</u>

Δίνεται παρατήρηση x από ομοιόμορφη $x|\vartheta\sim\mathcal{U}\left(\cdot|0,\vartheta\right)$, $\vartheta>0$ και gamma prior $\vartheta\sim Ga\left(\cdot|2,\beta\right)$.

1. Να βρεθεί η από κοινού κατανομή των x και ϑ.
2. Να δειχτεί ότι η prior predictive είναι εκθετική με παράμετρο β ενώ η posterior είναι εκθετική με παράμετρο β περικομμένη (truncated exponential) στο διάστημα $\left(x,\infty\right)$.
3. Να βρεθεί η εκτίμηση του ϑ ως προς τετραγωνική συνάρτηση απώλειας.
4. Να βρεθεί η posterior predictive για μια μελλοντική (unobserved) παρατήρηση $y|\vartheta\sim\mathcal{U}\left(\cdot|0,\vartheta\right)$.
5. Εάν $x_i|\vartheta\sim\mathcal{U}\left(\cdot|0,\vartheta\right)$, $i=1,\ldots,n$ ποία η εκλογή του $\pi\left(\vartheta\right)$ έτσι ώστε η $\pi\left(\vartheta\,|\,x\right)$ να είναι περικομμένη εκθετική ?
6. Να υπολογιστούν οι πυκνότητες $\pi\left(x,\vartheta\right)$, $\pi\left(x\right)$ και $\pi\left(\vartheta|x\right)$, όταν $x|\vartheta\sim\mathcal{U}\left(\cdot|0,\vartheta\right)$ και $\vartheta\sim Ga\left(\cdot|3,\beta\right)$.

1. $\pi\left(x|\vartheta\right)=\mathcal{U}\left(x\,|0,\vartheta\right)=\vartheta^{-1}\,1\left(0<x<\vartheta\right)$ και $\pi\left(\vartheta\right)=Ga\left(\vartheta\,|2,\beta\right)=\beta^{2}\vartheta e^{-\beta\vartheta}1\left(\vartheta>0\right)$,

η από κοινού κατανομή των x και ϑ είναι

$\pi\left(x,\vartheta\right)=\pi\left(\vartheta\right)\pi\left(x|\vartheta\right)=\beta^{2}\vartheta e^{-\beta\vartheta}\times\vartheta^{-1}1\left(0<x<\vartheta\right)=\beta^{2}e^{-\beta\vartheta}1\left(0<x<\vartheta\right)$.

2. Η prior predictive είναι $\pi\left(x\right)=\int_{\mathbb{R}^{+}}\pi\left(x,\vartheta\right)d\vartheta=\beta^{2}\int_{\mathbb{R}^{+}}e^{-\beta\vartheta}1\left(0<x<\vartheta\right)d\vartheta$

$=\beta^{2}\int_{x}^{\infty}e^{-\beta\vartheta}d\vartheta=\beta e^{-\beta x}=Exp\left(x\,|\,\beta\right)$.

Η posterior είναι $\pi(\vartheta|x)=\dfrac{\pi(x,\vartheta)}{\pi(x)}=\beta e^{-\beta(\vartheta-x)}1(\vartheta>x)$, για $x>0$.

Η posterior είναι περικομμένη εκθετική στο διάστημα (x,∞) εφόσον $\pi(\vartheta|x)\propto Exp(\vartheta\,|\,\beta)\cdot 1(\vartheta>x)$ είναι ισοδύναμο με το ότι $\pi(\vartheta|x)=C\beta e^{-\beta\vartheta}1(\vartheta>x)$ και $C^{-1}=\int_0^\infty \beta e^{-\beta\vartheta}1(\vartheta>x)\,d\vartheta\ =\beta\int_x^\infty e^{-\beta\vartheta}d\vartheta=e^{-\beta x}$ από όπου $\pi(\vartheta|x)=\beta e^{-\beta(\vartheta-x)}1(\vartheta>x)$.

Σημειώστε ότι: εάν $\Theta\sim Exp(\beta)$ τότε η υπό συνθήκη τ.μ. $[\Theta\,|\,\Theta>x]$ για $x>0$, είναι περικομμένη εκθετική στο διάστημα (x,∞).

$$Exp(\vartheta\,|\,\beta,\vartheta>x)\,d\vartheta=P\{\vartheta<\Theta\le\vartheta+d\vartheta\,|\,\Theta>x\}$$

$$=\frac{P\{\vartheta<\Theta\le\vartheta+d\vartheta,\Theta>x\}}{P\{\Theta>x\}}=\frac{P\{\vartheta<\Theta\le\vartheta+d\vartheta\}\,P\{\Theta>x\,|\,\vartheta<\Theta\le\vartheta+d\vartheta\}}{P\{\Theta>x\}}$$

$$=\frac{P\{\vartheta<\Theta\le\vartheta+d\vartheta\}1(\vartheta>x)}{P\{\Theta>x\}}=\frac{Exp(\vartheta\,|\,\beta)\,d\vartheta\,1(\vartheta>x)}{\int_{\vartheta=x}^\infty Exp(\vartheta\,|\,\beta)\,d\vartheta}=\beta e^{-\beta(\vartheta-x)}1(\vartheta>x).$$

3. $\hat{\vartheta}_{Bayes}=E(\vartheta|x)=\beta e^{\beta x}\int_{\mathbb{R}}\vartheta e^{-\beta\vartheta}1(\vartheta>x)\,d\vartheta=\beta e^{\beta x}\int_{\vartheta>x}\vartheta e^{-\beta\vartheta}d\vartheta$

$$=x+\frac{1}{\beta}=(\vartheta)_{MLE}+\frac{1}{2}\mathbb{E}(\vartheta).$$

4. $\pi(y\,|\,x)=\int_{\mathbb{R}}\mathcal{U}(y\,|\,0,\vartheta)\,\beta e^{-\beta(\vartheta-x)}1(\vartheta>x)\,d\vartheta$

$$=\int_{\vartheta>x}\vartheta^{-1}1(\vartheta>y)\,\beta e^{-\beta(\vartheta-x)}d\vartheta=\beta e^{\beta x}\int_{\vartheta>z}\vartheta^{-1}e^{-\beta\vartheta}d\vartheta,$$

όπου $z=\max\{x,y\}$. Κάνοντας την αλλαγή κλίμακας $\vartheta=zu$ έχουμε

$$\pi(y\,|\,x)=\beta e^{\beta x}\int_{u=1}^\infty u^{-1}e^{-\beta zu}du=\beta e^{\beta x}Ei(1,\beta z),$$

όπου $Ei(n,\alpha)=\int_{u=1}^{\infty}u^{-n}e^{-\alpha u}du$ είναι η συνάρτηση «εκθετικό ολοκλήρωμα» (the exponential integral function) , που ορίζεται για $n\in\mathbb{N}$ και $\alpha>0$.

5. Γνωρίζουμε ότι $\pi(x|\vartheta)\overset{\vartheta}{\propto}\vartheta^{-n}\,1\left(x_{(n)}\le\vartheta\right)$, θέτοντας $\pi(\vartheta)=Ga(\vartheta\,|\,a,\beta)\propto\vartheta^{a-1}e^{-\beta\vartheta}$, έχουμε $\pi(\vartheta|x)\propto\vartheta^{a-1}e^{-\beta\vartheta}\times\vartheta^{-n}1\left(\vartheta\ge x_{(n)}\right)$. Θέτοντας $a=n+1$, παίρνουμε $\pi(\vartheta|x)\propto e^{-\beta\vartheta}1\left(\vartheta\ge x_{(n)}\right)$ που είναι εκθετική με παράμετρο β, περικομμένη στο διάστημα $\left(x_{(n)},\infty\right)$. **Αυτή η επιλογή για prior δεν είναι τόσο καλή εφόσον η prior σε αυτή την περίπτωση εξαρτάται από το μέγεθος του δείγματος**.

6. Εδώ έχουμε ότι $\pi(\vartheta)=\frac{1}{2}\beta^3\vartheta^2e^{-\beta\vartheta}$ που δίνει

$$\pi(x,\vartheta)=\frac{1}{2}\beta^3\vartheta^2e^{-\beta\vartheta}\times\vartheta^{-1}1(0<x<\vartheta)=\frac{1}{2}\beta^3\vartheta e^{-\beta\vartheta}1(0<x<\vartheta).$$

Η prior predictive είναι $\pi(x)=\int_{\mathbb{R}^+}\pi(x,\vartheta)d\vartheta=\frac{\beta^3}{2}\int_{\vartheta>x}\vartheta e^{-\beta\vartheta}d\vartheta$

$$=\frac{1}{2}\beta^2xe^{-\beta x}+\frac{1}{2}\beta^2e^{-\beta x}=\frac{1}{2}Ga(x\,|\,2,\beta)+\frac{1}{2}Ga(x\,|\,1,\beta),$$

Η posterior τότε γίνεται

$$\pi(\vartheta|x)=\frac{\pi(x,\vartheta)}{\pi(x)}=\frac{\frac{1}{2}\beta^3\vartheta e^{-\beta\vartheta}1(\vartheta>x)}{\frac{1}{2}Ga(x\,|\,2,\beta)+\frac{1}{2}Ga(x\,|\,1,\beta)}=\frac{\beta^3\vartheta e^{-\beta\vartheta}1(\vartheta>x)}{Ga(x\,|\,2,\beta)+Ga(x\,|\,1,\beta)}$$

$$=\frac{\beta^3\vartheta e^{-\beta\vartheta}}{\beta^2xe^{-\beta x}+\beta e^{-\beta x}}1(\vartheta>x)=\frac{\beta^2}{1+\beta x}\vartheta e^{-\beta(\vartheta-x)}1(\vartheta>x),\ x>0[0,1],$$

που είναι μια $Ga(\vartheta\,|\,2,\beta)$ περικομμένη στο διάστημα (x,∞).

Πράγματι εάν για την πυκνότητα $f(\vartheta|x)$ ισχύει ότι $f(\vartheta|x) \propto Ga(\vartheta|2,\beta)\cdot 1(\vartheta > x)$, θα έχουμε για $C>0$ ότι $f(\vartheta|x) = C\cdot Ga(\vartheta|2,\beta)\cdot 1(\vartheta > x)$ που δίνει

$$C^{-1} = \int_{-\infty}^{\infty} Ga(\vartheta|2,\beta)\cdot 1(\vartheta > x)\,d\vartheta = \beta^2 \int_{\vartheta=x}^{\infty} \vartheta e^{-\beta\vartheta} d\vartheta = e^{-\beta x}(1+\beta x),$$

από όπου παίρνουμε $f(\vartheta|x) = \dfrac{\beta^2}{1+\beta x}\vartheta e^{-\beta(\vartheta - x)} 1(\vartheta > x)$.

<u>Άσκηση</u>

Το προηγούμενο παράδειγμα για n παρατηρήσεις x_i από ομοιόμορφη $x_i|\vartheta \sim \mathcal{U}(\cdot|0,\vartheta),\ \vartheta > 0$.

<u>Άσκηση</u>

Να δειχτεί ότι η συνάρτηση $\pi(y|x) = \beta e^{\beta x} Ei(1,\beta z)$, όπου $z = \max\{x,y\} = x \vee y$ με supp$(y) = (0,\infty)$, είναι πυκνότητα.

$$\int_{y>0} \pi(y|x)\,dy = \int_{y\le x} \pi(y|x)\,dy + \int_{y>x} \pi(y|x)\,dy$$

$$= \int_{0<y\le x} \beta e^{\beta x} Ei(1,\beta x)\,dy + \int_{y>x} \beta e^{\beta x} Ei(1,\beta y)\,dy$$

$$= \beta x e^{\beta x} Ei(1,\beta x) + \beta e^{\beta x} \int_{y>x} Ei(1,\beta y)\,dy$$

$$\int_{y=x}^{\infty} Ei(1,\beta y)\,dy = \int_{y=x}^{\infty}\int_{t=1}^{\infty} t^{-1} e^{-\beta y t}\,dt\,dy = \int_{t=1}^{\infty} t^{-1} \int_{y=x}^{\infty} e^{-\beta y t}\,dy\,dt$$

$$= \beta^{-1} \int_{t=1}^{\infty} t^{-2} e^{-\beta x t}\,dt = \beta^{-1} Ei(2,\beta x)$$

Δηλαδή έχουμε ότι $\int_{y>0} \pi(y|x)\,dy = e^{\beta x}\{\beta x Ei(1,\beta x) + Ei(2,\beta x)\}$,

όμως

$$Ei(1,u) = -u^{-1}\int_{t=1}^{\infty} t^{-1} d\left(e^{-ut}\right) = -u^{-1}\left\{-e^{-u} - \int_{t=1}^{\infty} e^{-ut}\left(-t^{-2}\right)dt\right\} = u^{-1}e^{-u} - u^{-1}Ei(2,u),$$

που τελικά δίνει:

$$\int_{y>0} \pi\left(y \mid x\right) dy = e^{\beta x}\left\{\beta x\left[\frac{e^{-\beta x}}{\beta x} - \frac{1}{\beta x} Ei\left(2, \beta x\right)\right] + Ei\left(2, \beta x\right)\right\} = 1.$$