Παρασκευή 13/11/20

Η επιλογή της εκ των προτέρων κατανομής είναι καθοριστικής σημασίας στη Μπεϋζιανή συμπερασματολογία, όμως μπορεί να είναι μια δύσκολη διαδικασία, καθώς δεν διαθέτουμε πάντα επαρκείς πληροφορίες ή οι διαθέσιμες πληροφορίες μπορεί να είναι συμβατές με περισσότερες από μια υποψήφιες εκ των προτέρων κατανομές.

Οι συνηθέστερες μέθοδοι επιλογής της εκ των προτέρων κατανομής είναι οι:

- συζυγείς εκ των προτέρων κατανομές
- υποκειμενικός καθορισμός των εκ των προτέρων κατανομών
- μη-πληροφοριακές εκ των προτέρων κατανομών.

Η ουσία είναι να βρούμε μια οικογένεια κατανομών, εξαρτώμενη από μια ή περισσότερες παραμέτρους στην οποία να ανήκει και η εκ των προτέρων και η εκ των υστέρων κατανομή της παραμέτρου $\theta \in \Theta$

Συζυγείς εκ των Προτέρων κατανομές

Μια οικογένεια κατανομών $\mathcal{F}$ με στήριγμα Θ λέμε ότι είναι συζυγής για τη συνάρτηση πιθανοφάνειας $f(x|\theta)$ αν για κάθε εκ των προτέρων κατανομή $p \in \mathcal{F}$, η αντίστοιχη εκ των υστέρων ανήκει επίσης στην $\mathcal{F}$

Δηλαδή η εκ των προτέρων και η εκ των υστέρων ανήκουν στην ίδια οικογένεια κατανομών.
Σε αυτή την περίπτωση οι παράμετροι της εκ των υστέρων αποτελούν αναβάθμιση τιμών της εκ των προτέρων μέσω του δείγματος.

Η $f(x|\theta)$ ως συνάρτηση της θ, μας ενδιαφέρει μόνο ως προς το μεταβλητό της μέρος (παραλείπουμε τη σταθερά) και μπορεί ισοδύναμα να εκφραστεί ως προς την αντίστοιχη επαρκή στ. συνάρτηση X.

Αντιπροσωπευτικά μοντέλα είναι τα

- $X_i \sim B(1,\theta)$, $i=1..n$, τότε $X=\sum_{i=1}^{n} X_i \sim B(n,\theta)$.
- $X_i \sim P(\theta)$ $i=1,..n$ τότε $X=\sum X_i \sim P(n\theta)$
- $X_i \sim Exp(\theta)$ τότε $X=\sum X_i \sim Gamma(n\theta)$
- $X_i \sim N(\theta,\sigma^2)$ $X=\sum X_i \sim N(\theta,\sigma^2/n)$

Για τα μοντέλα αυτά οι συζυγείς οικογένειες εκ των προτέρων κατανομών δίνονται παρακάτω, όπου $X=x$ είναι η τιμή της επαρκούς στατιστικής συνάρτησης για το δείγμα.

$f(x|\theta)$	$p(\theta)$	$p(\theta|x)$ ανάλογο
$B(n,\theta) \propto \theta^{x}(1-\theta)^{n-x}$	$Beta(a,b) \propto \theta^{a-1}(1-\theta)^{b-1}$	$Beta(a+x, b+n-x) \propto \theta^{a+x-1}(1-\theta)^{b+n-x-1}$
$Poisson(n\theta) \propto e^{-n\theta}\theta^{x}$	$Gamma(a,b) \propto \theta^{a-1}e^{-b\theta}$	$Gamma(a+x, b+n) \propto \theta^{a+x-1}e^{-(b+n)\theta}$
$Gamma(n,\theta)\ \theta^{n} \propto e^{-x\theta}$	$Gamma(a,b) \propto \theta^{a-1}e^{-b\theta}$	$Gamma(a+n, b+x) \propto \theta^{a+n-1}e^{-(b+x)\theta}$
$N(\theta,\sigma^2/n) \propto e^{-n(x-\theta)^2/2\sigma^2}$	$N(\mu,\tau^2) \propto$ $N(\mu,\tau^2) \propto e^{-(\theta-\mu)^2/2\tau^2}$	$N\left(\frac{n\tau^2 x+\sigma^2\mu}{n\tau^2+\sigma^2}, \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}\right)$

Υποκειμενικός ορισμός εκ των προτέρων κατανομής

Δεν θεωρούμε συγκεκριμένη κατανομή αλλά καθορίζουμε την εκ των προτέρων κατανομή με βάση την υποκειμενική μας άποψη για την σχετική πιθανότητα μεμονωμένων τιμών του παραμετρικού χώρου π.χ) αν $\Theta = (\Theta_1, \Theta_2)$ και θεωρήσουμε ότι $p(\Theta_2) = 3p(\Theta_1)$ τότε καταλήγουμε ότι $p(\Theta_1) = \frac{1}{4}$, $p(\Theta_2) = \frac{3}{4}$

αφού $p(\Theta_1) + p(\Theta_2) = 1$

Η προσωπική μας άποψη πρέπει να είναι συμβατή με την Θεωρία Πιθανοτήτων.

π.χ) έστω $\Theta(\Theta_1, \Theta_2, \Theta_3)$ και πιστεύουμε ότι
$p(\Theta_1) = 3p(\Theta_3)$, $p(\Theta_1) = 3p(\Theta_2)$, $p(\Theta_2) = 2p(\Theta_3)$

$$\frac{p(\Theta_1)}{p(\Theta_3)} = \frac{p(\Theta_1) \cdot p(\Theta_2)}{p(\Theta_2) \cdot p(\Theta_3)} = \frac{3p(\Theta_2) \cdot 2p(\Theta_3)}{p(\Theta_2) \cdot p(\Theta_3)} = 6$$

το οποίο δεν συμφωνεί με την αρχική μας γνώμη
$p(\Theta_1) = 2p(\Theta_3)$

Στην περίπτωση συνεχούς παραμετρικού χώρου Θ η $p(\theta)$ δηλώνει πυκνότητα πιθανότητα και ο λόγος $p(\Theta_1)/p(\Theta_3)$ καλείται σχετική πιθανοφάνεια για αυτό και αυτός ο τρόπος προσδιορισμού καλείται μέθοδος της σχετικής πιθανοφάνειας.

Αν το Θ είναι συνεχές και φραγμένο σύνολο τότε μπορούμε να εκφράσουμε τη διαμέριση του $\Theta = \Theta_1 \cup \Theta_2 \cup \ldots \cup \Theta_k$

Αν $p(\Theta_i) = P(\theta \in \Theta_i)$ $i = 1, \ldots k$ η εκ των προτέρων υποκειμενική πληροφορία μπορεί να εκφραστεί π.χ ως
$p(\Theta_1) = 2p(\Theta_2) = 2p(\Theta_3) \ldots$
ή γενικά $c_1 p(\Theta_1) = \ldots = c_k P(\Theta_k)$, $c_1, \ldots c_k \in \mathbb{R}$
και $p(\Theta_1) + \ldots + p(\Theta_k) = 1$

Συνήθως $p(\theta_1) = \left(\sum_{j=1}^{k} \frac{c_1}{c_j}\right)^{-1}$, $p(\theta_i) = \frac{c_1}{c_j} p(\theta_1)$, $i=2..k$.

Συνήθως το Θ είναι ένα φραγμένο διάστημα της πραγματικής ευθείας, άρα μια υποκειμενική εκ των προτέρων κατανομή μπορεί να εκφραστεί σε ένα ιστόγραμμα και γι αυτό η μέθοδος καλείται και μέθοδος του ιστογράμματος.

Μια άλλη υποκειμενική εκ των προτέρων κατανομή που χρησιμοποιείται πολύ συχνά είναι η εκ των προτέρων δεδομένης συναρτησιακής μορφής. Συγκεκριμένα, επιλέγουμε ως εκ των προτέρων κατανομή ένα μέλος κάποιας οικογένειας κατανομών η οποία εξαρτάται από μια ή περισσότερες παραμέτρους (κανονική, βήτα, γάμμα) και καθορίζουμε τις τιμές των παραμέτρων βάσει της εκ των προτέρων πληροφορίας που διαθέτουμε, η οποία συνήθως αφορά χαρακτηριστικά της ροπής.

Παράδειγμα:
Έστω $\Theta = \mathbb{R}$ και πιστεύουμε ότι το θ έχει μέση τιμή 120 και τυπική απόκλιση 25. Τότε η εκ των προτέρων κατανομή θα μπορούσε να ήταν η $N(120, 25^2)$. Αν το θ είναι το ποσοστό επιτυχίας μιας επέμβασης, τότε $\Theta = (0,1)$ και η κλασική επιλογή εκ των προτέρων είναι η $Beta(a,b)$ όπου τα a, b προσδιορίζονται από τις σχέσεις

$$\mu = \frac{a}{a+b}, \quad \delta^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

Μη πληροφοριακές εκ των προτέρων κατανομές

- δεν δίνουν πληροφορίες για την θ.
- χαρακτηριστικές τέτοιες εκ των προτέρων κατανομές είναι
  - → οι καταχρηστικές/γενικευμένες
  - → οι Jeffreys
  - → οι αναλλοίωτες

Καταχρηστική εκ των προτέρων κατανομή

Μια μη αρνητική συνάρτηση p ορισμένη στο Θ θα λέγεται καταχρηστική (ή γενικευμένη) εκ των προτέρων κατανομή αν $\int_{\Theta} p(\theta)d\theta = +\infty$

Τέτοιες κατανομές μπορούν να οριστούν αν μπορεί να οριστεί κατά η εκ των υστέρων κατανομή, το οποίο επιτυγχάνεται αν για το παρατηρηθέν X μπορεί να οριστεί η περιθώρια πυκνότητα m(X) δηλ. αν.

$$m(x) = \int_{\Theta} f(x|\theta)p(\theta)d\theta < +\infty$$

Άσκηση:

Έστω $X \sim B(n,p)$, $\theta \in \Theta = (0,1)$

επιβεβαιώστε ότι η εκ των προτέρων κατανομή $p(\theta) = \frac{1}{\theta(1-\theta)}$ $\theta \in \Theta$ είναι καταχρηστική και ότι η εκ των υστέρων κατανομή ορίζεται για $x \neq 0, n$.

$$\sum_{i=0}^{1} p(\theta_i) = p(0) + p(1) = +\infty$$

Εκ των προτέρων κατανομή του Jeffreys

Ο Jeffreys (1946) πρότεινε μια διαδικασία καθορισμού μη-πληροφοριακής εκ των προτέρων κατανομής, βασισμένη στο μέτρο ή τον πίνακα πληροφορίας του Fisher.

$$I(\theta) = E\left\{\frac{\partial}{\partial\theta} \log f(X|\theta)\right\}^2 = -E\left\{\frac{\partial^2}{\partial\theta^2} \log f(X|\theta)\right\}$$

Η εκ των προτέρων κατανομή του Jeffreys είναι

$p_J(\theta) \propto \sqrt{I(\theta)}$

παραμένει αναλλοίωτη σε 1-1 μετασχηματισμούς.

Παράδειγμα: Jeffreys prior Διωνυμικής κατανομής

Έστω $X_1 \ldots X_n$ τ.δ. από $B(1,\theta)$, $\theta \in \Theta = (0,1)$

Να βρεθεί η $p_J(\theta)$

Το δείγμα που παρατηρήσαμε είναι το $x = (x_1 \ldots x_n)$ και η παράμετρος ενδιαφέροντος η θ.

$$f(x|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}, \quad x \in (0,1)$$

$$\log f(x|\theta) = (\Sigma x_i)\log\theta + (n-\Sigma x_i)\log(1-\theta), \quad x \in (0,1)^n$$

$$\frac{\partial}{\partial\theta} f(x|\theta) = \frac{\Sigma x_i}{\theta} - \frac{(n-\Sigma x_i)}{1-\theta}, \quad (x_1, \ldots x_n) \in (0,1)^n$$

$$\frac{\partial^2}{\partial\theta^2} f(x|\theta) = -\frac{\Sigma x_i}{\theta^2} - \frac{n-\Sigma x_i}{(1-\theta)^2}, \quad (x_1, \ldots x_n) \in (0,1)^n$$

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2} f(x,\theta)\right) = -E\left\{-\frac{\Sigma x_i}{\theta^2} - \frac{n-\Sigma x_i}{(1-\theta)^2}\right\}$$

$$\frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}$$

Τελικά $p_J(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta(1-\theta)}} \propto \theta^{-1/2}(1-\theta)^{-1/2}$.

$0<\theta<1$ άρα Jeffreys prior είναι η $Beta(1/2, 1/2)$

- Γενικά η $p_J(\theta)$ μπορεί να είναι και καταχρηστική εκ των προτέρων κατανομή.
- Ορίζεται και για πολυδιάστατες παραμέτρους αλλά δεν συνηθίζεται να χρησιμοποιείται, στην περίπτωση αυτή καταφεύγουμε σε μη-πληροφοριακές υποθέτοντας ότι $\theta_1, \ldots, \theta_k$ ανεξάρτητες

Για αυτό όταν $\theta = (\theta_1, \ldots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$, $k \geq 1$ τότε λέμε ότι η εκ των προτέρων του Jeffreys ισούται με $\sqrt{I(\theta)}$, όπου $I(\theta)$ πίνακας πληροφορίας Fisher.

Αναλλοίωτες εκ των προτέρων:

- δύο σημαντικές κατηγορίες παραμέτρων, τις παραμέτρους θέσης και κλίμακας

Παράμετρος θέσης:

Μια παράμετρος θ καλείται παραμ. θέσης της κατανομής μιας τ.μ. X αν η πυκνότητα της X έχει τη μορφή $f(x|\theta) = f(x-\theta)$, δηλαδή είναι συνάρτηση της διαφοράς $x-\theta$.

Παράμετρος κλίμακος:

αν $f(x|\theta) = \theta^{-1} f(x|\theta)$ εξαρτάται απο τον λόγο $\frac{x}{\theta}$.

π.χ) Στην $N(\theta, \sigma^2)$ με σ γνωστό, θ είναι παράμετρος θέσης
ενω στην $N(0, \theta^2)$, $U(0, \theta)$, Gamma(a, θ)
α γνωστό, θ παράμετρος κλίμακας

▷ Αν θ παράμετρος θέσης και $\Theta = \mathbb{R}$, τοτε η μη
πληροφοριακή αναλλοίωτη ως προς μετασχηματισμό θέσης
και εκ των προτέρων κατανομή του θ είναι η $p(\theta) = 1$

▷ Αν θ παράμετρος κλίμακας και $\Theta = (0, +\infty)$
τοτε αναλλοίωτη ως προς μετασχηματισμό κλίμακας
εκ των προτέρων κατανομή του θ είναι $p(\theta) = \frac{1}{\theta}$.