Κεφάλαιο 1

Στατιστική και Πιθανότητες

Ας υποθέσουμε ότι αναφερόμαστε σε έναν (αντικειμενικό) χώρο πιθανότητας $(\Omega, \mathcal{F}, \mathbb{P})$. Ο χώρος πιθανότητας αφορά στα ενδεχόμενα που επηρρεάζουν την αποζημίωση των ασφαλιστηρίων συμβολαίων μίας εταιρίας Α. Οι τυχ. μεταβλητές των αποζημιώσεων $X : \Omega \to \mathbb{R}_+$ είναι $\mathbb{F}$-μετρήσιμες και ανάλογα με την ύπαρξη των ροπών

$$\mathbb{E}(X^p),$$

ανήκουν σε κάποιον από τους γνωστούς από τη Συναρτησιακή Ανάλυση χώρους $L^p(\Omega, \mathcal{F}, \mathbb{P})$. Ο δε πραγματικός αριθμός

$$v_X = \sup\{v \in (0, \infty) | \mathbb{E}(X^v) < +\infty,$$

ονομάζεται *δείκτης ροπής* μίας μεταβλητής αποζημίωσης. Αν οι αποζημιώσεις ενός χαρτοφυλακίου ασφαλιστικών συμβολαίων $(X_1, X_2, ..., X_n)$ είναι *ανεξάρτητες* και ακολουθούν μία κατανομή F, τότε τίθεται το ερώτημα πώς εξακριβώνουμε ότι όντως -με κάποιο στατιστικό σφάλμα- οι τ.μ. ακολουθούν αυτήν την κατανομή.

1.1 Λίγη Απαραμετρική Στατιστική

Το *ομογενές* χαρτοφυλάκιο $(X_1, X_2, ..., X_n)$ που υποδείξαμε, ικανοποιεί τις ιδιότητες του *τυχαίου δείγματος*. Θεωρούμε λοιπόν τυχαίο δείγμα από κατανομή F. Η *εμπειρική συνάρτηση κατανομής* του τυχαίου δείγματος $(X_1, Q_2, ..., X_n)$ είναι η

$$F_n(x) = \frac{1}{n} \sum_{i=1} I_{(-\infty, x]}(X_i).$$

Αναφέρουμε -χωρίς απόδειξη- τα παρακάτω Θεωρήματα και Προτάσεις που αφορούν στην εμπειρική συνάρτηση κατανομής (ε.συν.κ.) του παραπάνω τυχαίου δείγματος.

1. Θ.1 Η ε.συν.κ. $F_n(x)$ του παραπάνω τ.δ. από κατανομή F έχει την ιδιότητα ότι η τυχαία μεταβλητή $nF_n(x)|x = x_0$ ακολουθεί $B(n, F(x_0))$.

2. Θ.2 Η τιμή της ε.συν.κ. $F_n(x)$ στο x είναι *ισχυρά συνεπής εκτιμήτρια* της $F(x)$, δηλαδή

$$\mathbb{P}(\lim_{n\to+\infty} F_n(x) = F(x)) = 1.$$

3. Θ.3 (Glivenko-Cantelli) Η F_n συγκλίνει ομοιόμορφα στην F, δηλαδή

$$\mathbb{P}(\lim_{n\to+\infty} \sup_{x\in\mathbb{R}} | \int_{\mathbb{R}} |F_n(x) - F(x)|dx = 0) = 1.$$

1.1.1 Kolmogorov-Smirnov Test

Επομένως έχουμε τον έλεγχο της

1. H_0 : Το τυχαίο δείγμα $X_1, X_2, ..., X_n$, ακολουθεί την κατανομή F_0 κατά της *εναλλακτικής*
2. H_1 : Το τυχαίο δείγμα $X_1, X_2, ..., X_n$, ΔΕΝ ακολουθεί την κατανομή F_0.

Το K-S τεστ για τον έλεγχο αυτόν χρησιμοποιεί τη συνάρτηση

$$D_n = \sup_{x\in\mathbb{R}} |F_n(x) - F_0(x)|,$$

με περιοχή απόρριψης

$$\mathbf{K} : D_n > D_n(a),$$

δηλαδή

$$\mathbb{P}(D_n > D_n(a)|H_0) = a.$$

1.1.2 Μετασχηματισμός Ολοκληρώματος Πιθανότητας για το K-S

Προσεγγιστικοί τύποι για το κριτήριο D_n μπορούν να υπολογιστούν μέσω του μετασχηματισμού *ολοκληρώματος πιθανότητας*. Αν μία τ.μ. X ακολουθεί μία κατανομή F, τότε η $Y = F(X)$ ακολουθεί την ομοιόμορφη στο $(0, 1)$.
Απόδειξη.
Ορίζουμε την αντίστροφη της συνάρτηση κατανομής F ως εξής :

$$F^{-1}(y) = \inf\{x|F(x) \geq y\}, y \in (0, 1).$$

Από γνωστή θεωρία συναρτήσεων, $F(F^{-1}(y)) = y, y \in (0, 1)$. Εξ αιτίας της μονοτονίας της F, έχουμε ότι

$$\{X \leq F^{-1}(y)\} = \{F(X) \leq y\} = \{Y \leq y\}.$$

Άρα σύμφωνα με γνωστή Θεωρία Κατανομών, το δείγμα των *διατεταγμένων* παρατηρήσεων $X_{(i)}, i = 1, 2, ..., n$, έχει την ιδιότητα ότι τα δειγματικά σημεία

$$t_{(i)} = F(X_{(i)}), i = 1, 2, ..., n,$$

ακολουθούν την κατανομή Beta(i, n-i+1) δηλαδή η συν.π.π αυτής είναι

$$g_i(t) = \frac{n!}{(n-i)!(i-1)!} t^{i-1}(1-t)^{n-i}, i = 1, 2, ..., n,$$

όπου

$$\mathbb{E}(t_{(i)}) = \frac{i}{n+1}, V(t_{(i)}) = \frac{i(n+1-i)}{n+1)^2(n+2)}, i = 1, 2, ..., n.$$

Επομένως, το D_n προσεγγίζεται από $\max\{D_n^+, D_n^-\}$, όπου

$$D_n^+ = \max_{1 \le i \le n} \{\frac{i}{n} - t_{(i)}\}, D_n^- = \max_{1 \le i \le n} \{t_{(i)} - \frac{(i-1)}{n}\}.$$

Παρατηρούμε ότι όντως τα $D_n^{+,-}$ είναι *ανεξάρτητα από κατανομές*.

1.1.3 Ασκήσεις

1. Μελετήστε το απαραμερικό τεστ Anderson - Darling.

2. Σε ποια κατανομή προσαρμόζονται ικανοποιητικά τα νούμερα

   $$23, 12, 12, 13, 24, 12, 56, 79, 89, 29, 55, 33, 55, 66$$

   με $a = 0.95$;

3. Το ίδιο για τα νούμερα

   $$49, 59, 67, 89, 45, 45, 45, 67, 89, 100, 123, 123, 134, 129, 134, 156, 139$$

1.2 Βαριές Ουριές

Συνήθως, οι αποζημιώσεις στα ασφαλιστήρια συμβόλαια πληρώνονται χάρη σε ενδεχόμενα με *πολύ μικρή πιθανότητα*. Ανάλογα όμως με το πόσο καταστροφικό είναι το υπό ασφάλιση ατύχημα -ο κίνδυνος όπως λέμε, μία αποζημίωση ενδέχεται να είναι *πολύ μεγάλη*. Τυχαίες μεταβλητές που με μικρή πιθανότητα παίρνουν πολύ μεγάλες θετικές ή αρνητικές τιμές ονομάζονται τ.μ. *με βαριές ουρές*. Ο ορισμός των μεταβλητών αυτών έχει ως εξής.

1. Μία τ.μ. έχει βαριά θετική ουρά αν $\mathbb{E}(e^{r \cdot X^+}) = \infty$, για κάθε $r > 0$.

2. Μία τ.μ. έχει βαριά αρνητική ουρά αν $\mathbb{E}(e^{r \cdot X^-}) = \infty$, για κάθε $r > 0$.

3. Μία τ.μ. έχει βαριά ουρά (ή αλλιώς και θετική και αρνητική βαριά ουρά), αν $\mathbb{E}(e^{r \cdot |X|}) = \infty$, για κάθε $r > 0$.

Μία τ.μ. που δεν έχει βαριά ουρά (αρνητική, θετική ή και τα δύο) λέμε ότι έχει ελαφριά ουρά (αντίστοιχα αρνητική, θετική ή και τα δύο). •

Παράδειγμα 1.2.1 *Μια κανονική κατανομή έχει ελαφριά ουρά. Έστω $X : N(\mu, \sigma^2)$. Αφού η ροπογεννήτριά της υπάρχει, δηλαδή $\mathbb{E}(e^{tX}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$, έχει ελαφριά ουρά.*

Παράδειγμα 1.2.2 *Μια κατανομή Pareto έχει βαριά ουρά. Υπόδειξη: Σκεφτείτε αναλόγως.*

Πρόταση 1.2.3 *Οι τ.μ. με ελαφριές (θετικές) ουρές X έχουν $v_X = +\infty$.*

Απόδειξη.
Έστω $r_0 > 0$ πραγματικός με $\mathbb{E}(e^{r_0 X}) \leq M_0, M_0 > 0$. Τότε για κάθε $k \in \mathbb{N}$, έχουμε $\frac{r_0^k}{k!}\mathbb{E}(X^k) \leq M_0$.

Πρόταση 1.2.4 *Οι τ.μ. με βαριές (θετικές) ουρές X έχουν $v_X < \infty$.*

Από τις Πιθανότητες γνωρίζουμε την Ανισότητα Markov:

$$\mathbb{P}(X \geq \epsilon) \leq \frac{\mathbb{E}(X)}{\epsilon}.$$

Η πιθανότητα αυτή σχετίζεται άμεσα με το αντικείμενο της Αντασφάλισης, μιας και έχει σχέση με την πιθανότητα η συνολική αποζημίωση ενός -ομογενούσ- χαρτοφυλακίου ασφαλειών να ξεπεράσει σε δεδομένο χρονικό ορίζοντα T το αρχικό κεφάλαιο της ασφαλιστικής εταιρίας u_0 συν τα ασφάλιστρα που έχει εισπράξει εντός του χρονικού ορίζοντα $[0, T]$-αν υποθέσουμε ότι ο ρυθμός είσπραξης ασφαλίστων στη μονάδα του χρόνου είναι c. Δηλαδή

$$\epsilon = u_0 + cT.$$

Το γεγονός ότι αυτή η ανισότητα δεν είναι καθόλου χρήσιμη για την εκτίμηση της πιθανότητας

$$\mathbb{P}(\sum_{i=1}^{n} X_i > u_0 + cT),$$

βρίσκεται στο δείκτη ροπής της $X_i =^d X$, που μπορεί όντας πεπερασμένος να είναι τέτοιος ώστε $\mathbb{E}(X) = \infty$.