ΘΕΩΡΙΑ ΚΑΤΑΝΟΜΩΝ

ΦΡΟΝΤΙΣΤΗΡΙΟ #8

ΥΠΕΝΘΥΜΙΣΗ ΘΕΩΡΙΑΣ

Διατεταγμένα στατιστικά Έστω $X_1, \dots, X_n$ ανεξάρτητες συνεχείς τ.μ. με κοινή συνάρτηση πυκνότητας f και κοινή συνάρτηση κατανομής F. Ορίζουμε: $X_{(1)}$ = η μικρότερη των $X_1, \dots, X_n$, $X_{(2)}$ = η δεύτερη στη σειρά μικρότερη των $X_1, \dots, X_n$, $\dots$, $X_{(j)}$ = η j-οστή μικρότερη των $X_1, \dots, X_n$, $\dots$, $X_{(n)}$ = η μεγαλύτερη των $X_1, \dots, X_n$, δηλαδή ισχύει ότι: $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Οι τ.μ. $X_{(1)}, \dots, X_{(n)}$ είναι γνωστές ως διατεταγμένα στατιστικά (order statistics) που αντιστοιχούν στις τ.μ. $X_1, \dots, X_n$. Π.χ. σε έναν αγώνα ταχύτητας 100 μέτρων συμμετέχουν 5 δρομείς. Υποθέτουμε ότι οι χρόνοι που επιτυγχάνουν οι αθλητές περιγράφονται από τις ανεξάρτητες τ.μ. $X_1, \dots, X_5$ οι οποίες επιπλέον ακολουθούν την ίδια κατανομή. Η τ.μ. $X_{(1)} = \min(X_1, \dots, X_5)$ αντιστοιχεί στον καλύτερο χρόνο που επιτεύχθηκε στην κούρσα, οι τ.μ. $X_{(2)}, X_{(3)}$ αντιστοιχούν στους χρόνους των αθλητών που πήραν το ασημένιο και το χάλκινο μετάλλιο αντίστοιχα και η τ.μ. $X_{(2)} - X_{(1)}$ εκφράζει τη διαφορά στους χρόνους άφιξης των δύο πρώτων αθλητών στο τέρμα της κούρσας. Αν οι χρόνοι, σε δευτερόλεπτα, που πέτυχαν οι 5 αθλητές ήταν: $X_1 = 9.9$, $X_2 = 10.5$, $X_3 = 9.5$, $X_4 = 11$, $X_5 = 10.3$, οι αντίστοιχες τιμές των τ.μ. $X_{(1)}, \dots, X_{(5)}$ θα είναι: $X_{(1)} = 9.5$, $X_{(2)} = 9.9$, $X_{(3)} = 10.3$, $X_{(4)} = 10.5$, $X_{(5)} = 11$. Η από κοινού συνάρτηση πυκνότητας της n-διάστατης συνεχούς τ.μ. $(X_{(1)}, \dots, X_{(n)})$ δίνεται από:

$$f_{(X_{(1)},\dots,X_{(n)})}(x_1,\dots,x_n)=n!\,f(x_1)\cdots f(x_n),\quad x_1<x_2<\dots<x_n$$

Η παραπάνω σχέση μπορεί να εξηγηθεί ως εξής: Για να είναι ίση η τ.μ. $(X_{(1)},\dots,X_{(n)})$ με $(x_1,\dots,x_n)$ αρκεί η τ.μ. $(X_1,\dots,X_n)$ να είναι ίση με κάποια από τις $n!$ μεταθέσεις της n-άδας των τιμών $(x_1,\dots,x_n)$. Δοθέντος ότι η από κοινού συνάρτηση πυκνότητας της τ.μ. $(X_1,\dots,X_n)$ είναι ίση με $f(x_{i_1})\cdots f(x_{i_n})$ για οποιαδήποτε μετάθεση $x_{i_1},\dots,x_{i_n}$ των τιμών $x_1,\dots,x_n$, η παραπάνω σχέση προκύπτει άμεσα.

Η περιθώρια συνάρτηση πυκνότητας της j-οστης διατεταγμένης στατιστικής $X_{(j)}$ είναι:

$$f_{X_{(j)}}(x)=\frac{n!}{(n-j)!\,(j-1)!}[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)\qquad(1)$$

Ένας άμεσος συλλογισμός που μας οδηγεί στην παραπάνω σχέση είναι ο εξής: Για να είναι ίση με x η τ.μ. $X_{(j)}$ θα πρέπει $j-1$ από τις n τιμές των τ.μ. $X_1,\dots,X_n$ να είναι μικρότερες από x, $n-j$ τιμές να είναι μεγαλύτερες από x και ακριβώς μία τιμή να είναι ίση με x. Η συνάρτηση πυκνότητας οποιουδήποτε συνόλου τιμών των τ.μ. $X_1,\dots,X_n$ τέτοιου ώστε $j-1$ τιμές να είναι μικρότερες από x, $n-j$ τιμές να είναι όλες μεγαλύτερες από x και ακριβώς μία τιμή να είναι ίση με x, είναι $[F(x)]^{j-1}[1-F(x)]^{n-j}f(x)$. Επειδή υπάρχουν $\binom{n}{j-1,\,n-j,\,1}=\frac{n!}{(n-j)!\,(j-1)!\,1!}=\frac{n!}{(n-j)!\,(j-1)!}$

διαφορετικοί τρόποι ώστε ένα σύνολο n τιμών των τ.μ. $X_1,\dots,X_n$ να είναι τέτοιο ώστε $j-1$ τιμές του να είναι όλες μικρότερες από x, $n-j$ τιμές του να είναι όλες μεγαλύτερες από x και ακριβώς μία τιμή του να είναι ίση με x, καταλήγουμε στον τύπο (1). Χρησιμοποιώντας παρόμοιους συλλογισμούς με αυτούς με τους οποίους καταλήξαμε στη σχέση (1), προκύπτει ότι η από κοινού συνάρτηση πυκνότητας της δισδιάστατης τ.μ. $(X_{(i)}, X_{(j)})$, με $i<j$, δίνεται από τον τύπο:

$$f_{(X_{(i)},X_{(j)})}(x_i,x_j)=\frac{n!\,[F(x_i)]^{i-1}}{(i-1)!(j-i-1)!(n-j)!}[F(x_j)-F(x_i)]^{j-i-1}$$

$$[1-F(x_j)]^{n-j}f(x_i)f(x_j) \qquad (2)$$

<u>#1</u> (Κατανομή εύρους τ.δ.) Έστω n ανεξάρτητες και ισόνομες τ.μ. $X_1,\dots,X_n$. Η τ.μ. $R=X_{(n)}-X_{(1)}$ καλείται εύρος (range) των παρατηρούμενων τ.μ. Αν οι τ.μ. $X_1,\dots,X_n$ έχουν κοινή συνάρτηση κατανομής F και κοινή συνάρτηση πυκνότητας f, τότε η συνάρτηση κατανομής της τ.μ. R μπορεί να εξαχθεί χρησιμοποιώντας τη σχέση (2), ως εξής: Για $a\geq 0$, διαδοχικά έχουμε:

$$P\{R\leq a\}=P\{X_{(n)}-X_{(1)}\leq a\}=\iint\limits_{x_n-x_1\leq a} f_{(X_{(1)},X_{(n)})}(x_1,x_n)\,dx_1\,dx_n$$

$$=\int_{-\infty}^{\infty}\int_{x_1}^{x_1+a}\frac{n!}{(n-2)!}[F(x_n)-F(x_1)]^{n-2}f(x_1)f(x_n)\,dx_n\,dx_1$$

Κάνουμε την αντικατάσταση $y=F(x_n)-F(x_1)$, $dy=f(x_n)dx_n$ και έχουμε:

$$\int_{x_1}^{x_1+a}[F(x_n)-F(x_1)]^{n-2}f(x_n)dx_n=\int_0^{F(x_1+a)-F(x_1)}y^{n-2}dy$$

$$=\frac{1}{n-1}[F(x_1+a)-F(x_1)]^{n-1}.$$

Συνεπώς, $P(R\le a)=n\int_{-\infty}^{\infty}[F(x_1+a)-F(x_1)]^{n-1}f(x_1)dx_1$. Στην ειδική περίπτωση κατά την οποία οι τ.μ. $X_1,\dots,X_n$ είναι ομοιόμορφα κατανεμημένες στο διάστημα $(0,1)$, για $0<a<1$, από την τελευταία σχέση, προκύπτει ότι:

$$P(R\le a)=n\int_0^1[F(x_1+a)-F(x_1)]^{n-1}f(x_1)dx_1$$

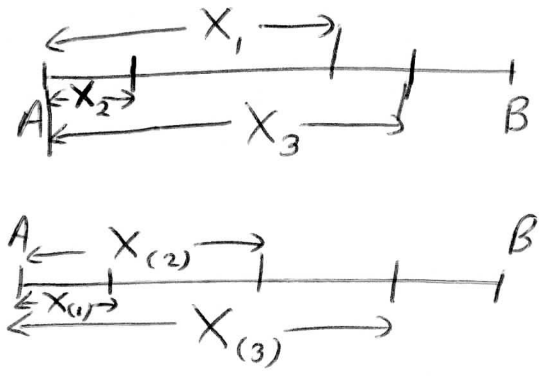
$$=n\int_0^{1-a}a^{n-1}dx_1+n\int_{1-a}^1(1-x_1)^{n-1}dx_1=n(1-a)a^{n-1}+a^n$$

Αν παραγωγίσουμε ως προς a, παίρνουμε τη συνάρτηση πυκνότητας του εύρους ως εξής:

$$f_R(a)=n(n-1)a^{n-2}(1-a),\ 0\le a\le 1 \text{ και } f_R(a)=0, \text{ αλλού.}$$

#2 Δύο πόλεις A και B απέχουν 30 Km και συνδέονται με έναν ευθύ αυτοκινητόδρομο (χωρίς καμία στροφή). Τρία άτομα επιλέγουν εντελώς τυχαία τη θέση όπου θα κτίσουν τα σπίτια τους, μεταξύ των δύο πόλεων και δίπλα στον αυτοκινητόδρομο. Ποια είναι η πιθανότητα η απόσταση μεταξύ οποιωνδήποτε διαδοχικών σπιτιών να είναι τουλάχιστον 10 Km;

Λύση



Έστω X_1, X_2, X_3 οι αποστάσεις των τριών σπιτιών από την πόρτα A. Τότε οι τ.μ. X_1, X_2, X_3 θα αποτελούν τ.δ. μεγέθους $n=3$ από την ομοιόμορφη κατανομή στο διάστημα $[0,30]$, δηλαδή

$$f(x)=\begin{cases}\frac{1}{30}, & 0\le x\le 30\\ 0, & \text{αλλού}\end{cases}$$

Η ζητούμενη πιθανότητα είναι ίση με:

$$p=P\left(X_{(2)}-X_{(1)}>10,\ X_{(3)}-X_{(2)}>10\right)$$

Η από κοινού συνάρτηση πυκνότητας των $X_{(1)}, X_{(2)}, X_{(3)}$ δίνεται από τον τύπο:

$$f_{123}(x,y,z)=3!\left(\frac{1}{30}\right)^3,\quad 0<x<y<z<30$$

και γράφουμε:

$$p=P\left(X_{(1)}+10<X_{(2)},\ X_{(2)}+10<X_{(3)}\right)$$

$$=\int_0^{10}\int_{x+10}^{20}\int_{y+10}^{30} f_{123}(x,y,z)\,dz\,dy\,dx$$

$$=3!\left(\frac{1}{30}\right)^3\int_0^{10}\int_{x+10}^{20}\int_{y+10}^{30} dz\,dy\,dx=\frac{1}{27}.\ \boxtimes$$

#3 Οι χρόνοι, σε sec, που πετυχαίνουν σε μία κούρσα 100 μέτρων οι 5 δρομείς που συμμετέχουν σ' αυτή, περιγράφονται από τις ανεξάρτητες τ.μ. $X_1, X_2, \ldots, X_n$, καθεμία από τις οποίες ακολουθεί Ομοιόμορφη κατανομή στο διάστημα (9,11). Να βρεθεί ο αναμενόμενος χρόνος του μεσαίου δρομέα της κούρσας, δηλαδή αυτού που τερματίζει τρίτος.

Λύση Μας ενδιαφέρει η μέση τιμή της τ.μ. $X_{(3)}$. Για τις συναρτήσεις πυκνότητας $f(x)$ και τις συναρτήσεις κατανομής των τ.μ. X_i έχουμε:

$$f(x) = \begin{cases} \frac{1}{2}, & 9 \le x \le 11 \\ 0, & \text{αλλού} \end{cases} \qquad F(x) = \begin{cases} 0, & x < 9 \\ \frac{x-9}{2}, & 9 \le x \le 11 \\ 1, & x > 11 \end{cases}$$

Άρα,

$$f_{X_{(3)}}(x) = \frac{5!}{2!\,3!} f(x) [F(x)]^2 [1-F(x)]^2$$

$$= 10 (x-9)^2 (11-x)^2, \quad 9 \le x \le 11$$

Η ζητούμενη μέση τιμή είναι ίση με:

$$E[X_{(3)}] = \int_{-\infty}^{\infty} x f_{X_{(3)}}(x)\,dx = 10 \int_9^{11} x (x-9)^2 (11-x)^2 dx = \frac{320}{3}.$$

#5 Ένας αυτόματος μηχανισμός παράγει ήχους των οποίων η ένταση ακολουθεί μία συνεχή κατανομή με συνάρτηση πυκνότητας $f(x) = 2x$, $0 < x < 1$. Αν παραχθούν τέσσερις τέτοιοι ήχοι, ποια είναι η πιθανότητα, η υψηλότερη ένταση που θα παρατηρηθεί να είναι τουλάχιστον διπλάσια της χαμηλότερης;

Λύση Αν συμβολίσουμε με X_1, X_2, X_3, X_4 τις εντάσεις των ήχων που θα παραχθούν τότε η ζητούμενη πιθανότητα είναι η $P(X_{(4)} > 2X_{(1)})$. Οι τ.μ. X_1, X_2, X_3, X_4 αποτελούν τ.δ. μεγέθους $n=4$ από μία συνεχή κατανομή με συνάρτηση πυκνότητας $f(x)=2x$, $0<x<1$ και συνάρτηση κατανομής $F(x)=\int_0^x 2t\,dt = x^2$, $0<x<1$.
Η από κοινού συνάρτηση πυκνότητας των τ.μ. $X_{(1)}, X_{(4)}$ είναι:

$$f_{(X_{(1)},X_{(4)})}(x,y) = \frac{4!}{0!\,2!\,0!} f(x) f(y) [F(x)]^0 [F(y)-F(x)]^2 [1-F(y)]^0$$

$$= 48xy(y^2-x^2)^2, \quad 0<x<y<1$$

και

$$P(X_{(4)} > 2X_{(1)}) = \int_0^1 \int_0^{y/2} 48xy(x^2-y^2)^2\,dx\,dy$$

$$= \int_0^1 \frac{37}{8} y^7\,dy = \frac{37}{64}.$$

#6 Τρεις φοιτητές δίνουν ραντεβού για φαγητό μεταξύ 12:00 και 13:00, σε ένα εστιατόριο. Αν οι χρόνοι άφιξης των τριών φοιτητών στο εστιατόριο είναι τυχαίοι, να υπολογιστεί η μέση τιμή του χρόνου που απαιτείται για να συγκεντρωθούν και οι τρεις φοιτητές στο εστιατόριο.

Λύση Έστω X_1, X_2, X_3 οι χρόνοι άφιξης των τριών φοιτητών στο εστιατόριο. Οι τ.μ. X_1, X_2, X_3 αποτελούν τ.δ. μεγέθους $n=3$ από την Ομοιόμορφη κατανομή στο διάστημα $(0,1)$ η οποία έχει συνάρτηση πυκνότητας και συνάρτηση κατανομής: $f(x)=1$, $F(x)=x$, $0\leq x\leq 1$, αντίστοιχα.

Τότε $E[X_{(3)}-X_{(1)}]$ είναι η ζητούμενη μέση τιμή.

Είναι

$$f_{(X_{(1)},X_{(3)})}(x,y)=\frac{3!}{0!\,1!\,0!}f(x)f(y)[F(x)]^0[F(y)-F(x)]^1[1-F(y)]^0$$

$$=6(y-x),\quad 0\leq x<y\leq 1$$

και

$$E[X_{(3)}-X_{(1)}]=\int_0^1\int_x^1 6(y-x)^2\,dy\,dx=$$

$$=\int_0^1(-2x^3+6x^2-6x+2)\,dx=\frac{1}{2}.\ \boxtimes$$