ΦΡΟΝΤΙΣΤΗΡΙΟ #2

Συνέχεια στις εφαρμογές του Κ.Ο.Θ.

#1 Το πλήθος των αριθμών προτεραιότητας που δίνει σε μία ημέρα ένα αυτόματο μηχάνημα στην αίθουσα αναμονής ενός εξωτερικού ιατρείου ακολουθεί την κατανομή Poisson με παράμετρο $\lambda=50$. Ποια είναι προσεγγιστικά η πιθανότητα:
(i) να δοθούν σε μία μέρα από 35 έως 70 αριθμοί;
(ii) το πλήθος των αριθμών που θα δοθούν εντός ενός έτους (200 εργάσιμες μέρες) να είναι μεταξύ 9950 και 11000;

Λύση (i) Έστω X τ.μ. που αναπαριστά το πλήθος των αριθμών προτεραιότητας που δίνει το αυτόματο μηχάνημα σε μία μέρα. Αφού X ακολουθεί την κατανομή Poisson με παράμετρο η οποία έχει αρκετά μεγάλη τιμή*, μπορούμε να υπολογίσουμε τη ζητούμενη πιθανότητα μέσω της κανονικής προσέγγισης υποθέτοντας ότι η τ.μ.

$$Z=\frac{X-E[X]}{\sqrt{var(X)}}=\frac{X-50}{\sqrt{50}}\sim N(0,1)$$

(*) Αν η παράμετρος λ παίρνει μεγάλες τιμές μπορούμε να θεωρήσουμε ότι γράφεται σε μορφή αθροίσματος πολλών προσθετέων $\lambda=\lambda_1+\lambda_2+\dots+\lambda_n$, καθένας από τους οποίους δεν είναι αμελητέος. Αν θεωρήσουμε n ανεξάρτητες τ.μ. $X_1,\dots,X_n$ τ.ω. η X_i, $i=1,\dots,n$ να ακολουθεί την κατανομή Poisson με παράμετρο λ_i, $i=1,\dots,n$, το άθροισμα $S_n=X_1+X_2+\dots+X_n$ θα ακολουθεί την κατανομή Poisson με παράμετρο

$\lambda_1+\lambda_2+\dots+\lambda_n=\lambda$. Αφού το πλήθος των προσθετέων είναι μεγάλο, η τ.μ. $\dfrac{S_n-E(S_n)}{\sqrt{var(S_n)}}$ θα ακολουθεί, κατά προσέγγιση, την κανονική κατανομή $N(0,1)$. Επιπλέον, $E(S_n)=var(S_n)=\sum_{i=1}^{n}\lambda_i=\lambda$, και η τ.μ. S_n έχει την ίδια ακριβώς κατανομή με την X.

Για τη ζητούμενη πιθανότητα, έχουμε:

$$P(35\le X\le 70)=P\left(\frac{35-50}{\sqrt{50}}\le\frac{X-50}{\sqrt{50}}\le\frac{70-50}{\sqrt{50}}\right)$$

$$=P(-2.12\le Z\le 2.83)\underset{\text{Κ.Ο.Θ.}}{\cong}\Phi(2.83)-\Phi(-2.12)$$

$$\cong\Phi(2.83)-(1-\Phi(2.12))=0.9807$$

Με διόρθωση συνέχειας

$$P(35\le X\le 70)\cong\Phi\left(\frac{70+0.5-50}{\sqrt{50}}\right)-\Phi\left(\frac{35-0.5-50}{\sqrt{50}}\right)$$

$$=\Phi(2.90)-\Phi(-2.19)=\Phi(2.90)-(1-\Phi(2.19))=$$

$$=0.9838$$

Η ακριβής τιμή της ζητούμενης πιθανότητας είναι:

$$P(35\le X\le 70)=\sum_{x=35}^{70}e^{-50}\frac{50^x}{x!}=0.9862$$

(με χρήση υπολογιστή).

(ii) Έστω X_i το πλήθος των αριθμών προτεραιότητας που δίνει το αυτόματο μηχάνημα την i-οστή εργάσιμη μέρα του έτους, $i=1,\ldots,200$. Το πλήθος των αριθμών που θα δοθούν σε ένα έτος θα εκφράζεται από το άθροισμα $S'_{200} = X_1 + X_2 + \ldots + X_{200}$.

Είναι: $E(S_{200}) = 200 \cdot 50 = 1000$

$\mathrm{var}(S_{200}) = 200 \cdot 50 = 1000$

Σύμφωνα με το Κ.Ο.Θ., η τυποποιημένη τ.μ.

$$Z = \frac{S'_{200} - E(S'_{200})}{\sqrt{\mathrm{var}(S_{200})}} = \frac{S' - 1000}{\sqrt{1000}} \underset{\text{προσεγγιστικά}}{\sim} N(0,1)$$

Επομένως,

$$P(950 \leq S'_{200} \leq 1.100)$$

$$= P\left(\frac{950-1000}{\sqrt{1000}} \leq \frac{S'-1000}{\sqrt{1000}} \leq \frac{1100-1000}{\sqrt{1000}}\right)$$

$$= P(-1.58 \leq Z \leq 3.16) \cong \Phi(3.16) - \Phi(-1.58)$$

$$= \Phi(3.16) - (1 - \Phi(1.58)) = 0.9421$$

Με διόρθωση συνέχειας:

$$P(950 \leq S'_{200} \leq 1100) \cong$$

$$\cong \Phi\left(\frac{1100+0.5-1000}{\sqrt{1000}}\right) - \Phi\left(\frac{950-0.5-1000}{\sqrt{1000}}\right)$$

$= \Phi(3.17) - \Phi(-1.60) = \Phi(3.17) - (1 - \Phi(1.60)) =$

$= 0.8881.$

#2 Έστω μ η άγνωστη πραγματική τιμή του pH μιας χημικής ουσίας. Χρησιμοποιώντας ένα όργανο, μετράμε την τιμή pH της ουσίας με n ανεξάρτητες μετρήσεις. Ας υποθέσουμε ότι κάθε μέτρηση είναι τ.μ. με μέση τιμή μ και διακύμανση σ^2. Πόσες μετρήσεις πρέπει να γίνουν ώστε με πιθανότητα τουλάχιστον α $(0<\alpha<1)$ η μέση τιμή των μετρήσεων να διαφέρει κατά απόλυτη τιμή από το μ το πολύ κατά $\sigma/_K$; Να γίνει εφαρμογή για $K=4,5,10$ και $\alpha = 90\%, 95\%$ και 99%.

Λύση Έστω X_i η τιμή που προέκυψε κατά την i-οστή μέτρηση, $i=1,2,\ldots,n$. Οι τ.μ. X_i είναι ανεξάρτητες και ισόνομες με $E[X_i]=\mu$, $Var[X_i]=\sigma^2$, οπότε, αν το n παίρνει μεγάλες τιμές, η τυποποιημένη τ.μ. $Z = \dfrac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{var(\bar{X}_n)}}$, όπου $\bar{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$, ακολουθεί, κατά προσέγγιση, την κανονική κατανομή $N(0,1)$.

Η συνθήκη είναι: $P\left(|\bar{X}_n - \mu| \le \frac{\sigma}{K}\right) \ge \alpha$

ή ισοδύναμα (αφού $E[\bar{X}_n]=\mu$, $Var(\bar{X}_n) = \frac{\sigma^2}{n}$)

$$P\left(\left|\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right| \le \frac{\sigma/k}{\sigma/\sqrt{n}}\right) \ge \alpha$$

δηλαδή $P\left(|Z_1| \le \frac{\sqrt{n}}{k}\right) \ge \alpha$

Αφού, κατά προσέγγιση, $Z_1 \sim N(0,1)$, βρίσκουμε

$$P\left(|Z_1| \le \frac{\sqrt{n}}{k}\right) = P\left(-\frac{\sqrt{n}}{k} \le Z_1 \le \frac{\sqrt{n}}{k}\right)$$

$$= \Phi\left(\frac{\sqrt{n}}{k}\right) - \Phi\left(-\frac{\sqrt{n}}{k}\right)$$

$$= \Phi\left(\frac{\sqrt{n}}{k}\right) - \left(1 - \Phi\left(\frac{\sqrt{n}}{k}\right)\right) = 2\Phi\left(\frac{\sqrt{n}}{k}\right) - 1$$

Επομένως

$$2\Phi\left(\frac{\sqrt{n}}{k}\right) \ge \alpha+1 \Rightarrow \Phi\left(\frac{\sqrt{n}}{k}\right) \ge \frac{\alpha+1}{2}$$

Για δεδομένα k και α, μπορούμε να προσδιορίσουμε την ελάχιστη τιμή του n.

Για παράδειγμα, για $k=4$, $\alpha=0.90$ πρέπει:

$$\Phi\left(\frac{\sqrt{n}}{4}\right) \ge 0.95$$

Αφού $\Phi(1.64) = 0.9495$

$\Phi(1.65) = 0.9505$

αρκεί $\Phi\left(\frac{\sqrt{n}}{4}\right) \ge \Phi(1.65)$ ή $\frac{\sqrt{n}}{4} \ge 1.65$

δηλαδή $n \ge [4 \cdot 1.65]^2 = 43.56$

Επομένως, η ελάχιστη τιμή του n είναι $n=44$. Στον πίνακα, δίνονται οι αντίστοιχες τιμές του n για τους υπόλοιπους συνδυασμούς των k και α.

$k \backslash \alpha$	0.90	0.95	0.99
4	44	62	70
5	69	97	167
10	173	385	666

Οι μεγάλες τιμές του n, δικαιολογούν τη χρήση του Κ.Ο.Θ.

#3 Δείξτε ότι $e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} \xrightarrow[n\to\infty]{} \frac{1}{2}$.

Λύση Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τ.μ. που ακολουθούν την κατανομή Poisson με παράμετρο ίση με τη μονάδα.

Η τ.μ. $n\overline{X}_n = n\frac{\sum_{i=1}^{n}X_i}{n} = \sum_{i=1}^{n}X_i$ ακολουθεί την κατανομή Poisson με παράμετρο n. Χρησιμοποιώντας το Κ.Ο.Θ. διαδοχικά έχουμε:

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = P\left(n\overline{X}_n \le n\right) = P\left(\overline{X}_n \le 1\right)$$

$$= P\left[\frac{\sqrt{n}\left(\overline{X}_n - E[X_i]\right)}{\sqrt{var(X_i)}} \le \frac{\sqrt{n}(1-1)}{\sqrt{1}}\right]$$

$$= P\left[\frac{\sqrt{n}\left(\overline{X}_n - E[X_i]\right)}{\sqrt{var(X_i)}} \le 0\right],\ i=1,\ldots,n \xrightarrow[\substack{n\to\infty \\ \text{Κ.Ο.Θ.}}]{} \Phi(0)=\frac{1}{2},$$

όπου Φ είναι η συνάρτηση κατανομής της $N(0,1)$.

#4 Ένας αστρονόμος ενδιαφέρεται να μετρήσει την απόσταση (σε έτη φωτός) του παρατηρητηρίου του από ένα μακρινό αστέρι. Παρόλο που ο αστρονόμος διαθέτει μία τεχνική μέτρησης, γνωρίζει ότι, λόγω των συνεχόμενων αλλαγών στις ατμοσφαιρικές συνθήκες και των "λαθών" στις μετρήσεις, η μέτρηση που κάνει κάθε φορά δεν αποδίδει την ακριβή απόσταση αλλά κάποια εκτίμησή της. Σχεδιάζει λοιπόν να κάνει μία σειρά από μετρήσεις και να χρησιμοποιήσει το μέσο όρο αυτών των μετρήσεων ως μία εκτίμηση για την πραγματική απόσταση. Αν ο αστρονόμος υποθέτει ότι οι τιμές των μετρήσεων είναι ανεξάρτητες και ισόνομες τ.μ. με κοινή μέση τιμή d (την πραγματική απόσταση) και κοινή διασπορά 4 ετών φωτός, πόσες μετρήσεις θα πρέπει να γίνουν ώστε ο αστρονόμος να "μοιάζει σίγουρος" ότι η εκτίμηση της απόστασης που υπολόγισε είναι ακριβής με απόκλιση ± 0.5 έτη φωτός;

Λύση Υποθέτουμε ότι ο αστρονόμος αποφασίζει να κάνει n μετρήσεις. Έστω $X_1, X_2, \dots, X_n$ οι τ.μ. που αναπαριστούν τις n μετρήσεις. Τότε, από το Κ.Ο.Θ. έχουμε:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nd}{2\sqrt{n}} \underset{\text{προσεγγιστικά}}{\sim} N(0,1)$$

Διαδοχικά, έχουμε:

$$P\left\{-0.5 \le \frac{\sum_{i=1}^{n} X_i}{n} - d \le 0.5\right\}$$

$$= P\left\{-0.5\frac{\sqrt{n}}{2} \le Z_n \le 0.5\frac{\sqrt{n}}{n}\right\}$$

$$\underset{\uparrow}{\cong} \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

Κ.Ο.Θ.

Επομένως, αν ο αστρονόμος θέλει, για παράδειγμα, να είναι 95% σίγουρος ότι η εκτίμηση που υπολόγισε είναι ακριβής με απόκλιση ± 0.5 έτη φωτός, θα πρέπει να κάνει n^* μετρήσεις, όπου n^* είναι τ.ω.

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = 0.95 \quad \text{ή} \quad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = 0.975$$

Ισοδύναμα, θα πρέπει: $\dfrac{\sqrt{n^*}}{4} = 1.96 \Rightarrow n^* = (7.84)^2 = 61.47$

Επομένως, θα πρέπει να κάνει 62 μετρήσεις.

Η παραπάνω ανάλυση βασίστηκε στην υπόθεση της κανονικής προσέγγισης (με την προϋπόθεση ότι η κανονική προσέγγιση είναι "καλή" προσέγγιση για μεγάλα n). Η απάντηση στο ερώτημα, "πόσο μεγάλο n θα μπορούσε να χρησιμοποιηθεί;" μπορεί να δοθεί μέσω της ανισότητας Chebyshev. Η ανισότητα δίνει έναν εναλλακτικό τρόπο για τον υπολογισμό του n.

Επειδή $E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = d$ και $var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{4}{n}$

η ανισότητα Chebyshev(*) δίνει:

$$P\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n} - d\right| > 0.5\right\} \leq \frac{4}{n(0.5)^2} = \frac{16}{n}$$

(*) Αν X τ.μ. με μέση τιμή μ και διασπορά σ^2, τότε για κάθε $k>0$, $P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$.

Αν απαιτήσουμε $P\left\{\left|\frac{\sum_{i=1}^{n} X_i}{n} - d\right| > 0.5\right\} = 0.05$

ο αστρονόμος θα πρέπει να κάνει $n = \frac{16}{0.05} = 320$ μετρήσεις ώστε να είναι 95% σίγουρος ότι η εκτίμησή του θα είναι ακριβής με απόκλιση 0.5 έτη φωτός. Γενικά, η ακρίβεια του υπολογισμού του n και της αξιοπιστίας της προσέγγισης με τη χρήση της ανισότητας Chebyshev, εξαρτάται από την κατανομή της τ.μ. X.