

HY-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες - Εαρινό Εξάμηνο 2011
Διδάσκων: Π. Τσακαλίδης

Λύσεις Πρώτης Σειράς Ασκήσεων

Ημερομηνία Ανάθεσης: 3/3/2011

Ημερομηνία Παράδοσης: 16/3/2011

Άσκηση 1.

(α) Έστω X_i τυχαία μεταβλητή που χαρακτηρίζει την ποιότητα του i -οστού επεξεργαστή (“0” για ελαττωματικό επεξεργαστή, “1” για μη ελαττωματικό επεξεργαστή). Τότε οι X_i με $i = 1, \dots, n$ είναι ανεξάρτητες Bernoulli τυχαίες μεταβλητές. Ορίζουμε την τ.μ. Z_n , ως:

$$Z_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Χρησιμοποιώντας την ανισότητα του Chebyshev έχουμε:

$$P(|Z_n - p| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2},$$

όπου σ^2 η διακύμανση της Bernoulli τυχαίας μεταβλητής. Έτσι, έχουμε ότι:

$$\lim_{n \rightarrow \infty} P(|Z_n - p| \geq \epsilon) = 0,$$

παρατηρώντας ότι $\lim_{n \rightarrow \infty} \frac{\sigma^2}{n\epsilon^2} = 0$. Αυτό σημαίνει ότι η Z_n συγκλίνει στο p ως προς την πιθανότητα.

(β) Χρήση ανισότητας Chebyshev:

Κάνοντας χρήση της ανισότητας Chebyshev, μπορούμε να δείξουμε ότι πρέπει να τεστάρουμε περισσότερους από 500 μικροεπεξεργαστές: Εφόσον η διασπορά σ^2 μιας Bernoulli τυχαίας μεταβλητής είναι $p(1-p)$, ισχύει ότι $\sigma^2 \leq \frac{1}{4}$. Συνεπώς, για $n \geq 500$ έχουμε:

$$\begin{aligned} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - p\right| \geq 0.1\right) &\leq \frac{\sigma^2}{n0.1^2} \\ &\leq \frac{\frac{1}{4}}{n0.1^2} \\ &\leq \frac{1}{4n} = 0.05. \end{aligned}$$

Παρ’ όλα αυτά, χρησιμοποιώντας την ανισότητα του Chebyshev δεν μπορούμε να πούμε αν λιγότεροι από 500 μικροεπεξεργαστές επαρκούν για το τεστ, μιας και δεν γνωρίζουμε τη διασπορά. Αν η διασπορά είναι πολύ μικρή, το οποίο μπορεί να συμβεί όταν το p είναι αρκετά μικρό, 27 μικροεπεξεργαστές θα μπορούσαν να επαρκούν στην πράξη. Συνεπώς, η απάντηση στο ερώτημα είναι “εξαρτάται”.

Χρήση κεντρικού οριακού θεωρήματος

Γνωρίζουμε από το ΚΟΘ ότι καθώς $n \rightarrow \infty$, η $Z_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ συμπεριφέρεται σαν κανονικά κατανοημένη τ.μ. με μέση τιμή p και διασπορά σ^2/n . Για πεπερασμένο n , αυτό δεν είναι ακριβές, αλλά μπορούμε να χρησιμοποιήσουμε το ΚΟΘ για να προσεγγίσουμε την CDF της Z_n . Η ποιότητα της προσέγγισης βελτιώνεται καθώς αυξάνεται το n .

Θέλουμε να βρούμε n τέτοιο ώστε $P(p - 0.1 \leq Z_n \leq p + 0.1) \geq 0.95$.

$$\begin{aligned} P(p - 0.1 \leq Z_n \leq p + 0.1) &= P\left(\frac{-0.1}{\sigma/\sqrt{n}} \leq \frac{Z_n - p}{\sigma/\sqrt{n}} \leq \frac{0.1}{\sigma/\sqrt{n}}\right) \\ &\approx \Phi\left(\frac{0.1}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-0.1}{\sigma/\sqrt{n}}\right) \\ &= 2\Phi\left(\frac{0.1}{\sigma/\sqrt{n}}\right) - 1. \end{aligned}$$

Στη χειρότερη περίπτωση η διασπορά της X_i είναι μεγάλη, οπότε απαιτούνται περισσότερα δείγματα n για να ικανοποιήσουμε το κριτήριο. Μέγιστη διασπορά έχουμε όταν $p = 1/2$, οπότε το σ^2 είναι $p(1-p) = 1/4$.

Συνεπώς, στη χειρότερη περίπτωση, $P(p - 0.1 \leq Z_n \leq p + 0.1) \approx 2\Phi(0.2\sqrt{n}) - 1$. Λύνοντας την εξίσωση $2\Phi(0.2\sqrt{n}) - 1 = 0.95$ ως προς n , έχουμε ότι $0.2\sqrt{n} = 1.96$ και προκύπτει ότι απαιτούνται $n \geq 96$ προσπάθειες.

Άσκηση 2.

(α) Εστω ότι οι τυχαίες μεταβλητές *Bernoulli* $X_1, X_2, \dots, X_{1000}$ έχουν κοινή συνάρτηση πιθανότητας

$$X_i = \begin{cases} 1, \text{ με πιθανότητα } 10^{-10} \\ 0, \text{ με πιθανότητα } 1 - 10^{-10} \end{cases}$$

Επομένως $E(X_i) = 10^{-10}$. Αν N είναι το πλήθος των λαθών σε 1000 *bits*, τότε

$$E(N) = E\left(\sum_{i=1}^{1000} X_i\right) = \sum_{i=1}^{1000} E(X_i) = 1000E(X_i) = 1000 \times 10^{-10} = 10^{-7}.$$

(β) Το άνω όριο μπορεί να βρεθεί από την ανισότητα *Marcov*:

$$P(N \geq \alpha) \leq \frac{E[N]}{\alpha},$$

όπου $\alpha = 10$ και $E[N] = 10^{-7}$. Επομένως $P(N \geq 10) \leq 10^{-8}$.

Άσκηση 3.

Χρησιμοποιώντας την De Moivre-Laplace προσέγγιση της διωνυμικής έχουμε:

(α)

$$\begin{aligned} P(190 \leq L \leq 210) &\approx \Phi\left(\frac{210 + \frac{1}{2} - 400 \cdot 0.51}{\sqrt{400 \cdot 0.51 \cdot 0.49}}\right) - \Phi\left(\frac{190 - \frac{1}{2} - 400 \cdot 0.51}{\sqrt{400 \cdot 0.51 \cdot 0.49}}\right) \\ &\approx \Phi(0.65) - \Phi(-1.45) \approx 0.6687 \end{aligned}$$

(β)

$$\begin{aligned} P(210 \leq L \leq 230) &\approx \Phi\left(\frac{230 + \frac{1}{2} - 400 \cdot 0.51}{\sqrt{400 \cdot 0.51 \cdot 0.49}}\right) - \Phi\left(\frac{210 - \frac{1}{2} - 400 \cdot 0.51}{\sqrt{400 \cdot 0.51 \cdot 0.49}}\right) \\ &\approx \Phi(2.65) - \Phi(0.55) \approx 0.2872 \end{aligned}$$

Άσκηση 4.

(α) Οι τυχαίες μεταβλητές Y_i δεν είναι ανεξάρτητες. Μπορούμε να το συμπεράνουμε διαισθητικά παρατηρώντας ότι διαδοχικές μεταβλητές Y_i εξαρτώνται από την ίδια τιμή της X_i . Μπορούμε, επίσης, να το αποδείξουμε κάνοντας χρήση της συνδιασποράς.

$$\text{Εφόσον } E[Y_i] = \frac{1}{3}E[X_i] + \frac{2}{3}E[X_{i+1}] = \mu,$$

$$\begin{aligned} \text{cov}(Y_i, Y_{i+1}) &= E[(Y_i - E[Y_i])(Y_{i+1} - E[Y_{i+1}])] \\ &= E\left[\left(\frac{1}{3}X_i + \frac{2}{3}X_{i+1} - \mu\right)\left(\frac{1}{3}X_{i+1} + \frac{2}{3}X_{i+2} - \mu\right)\right] \\ &= E\left[\left(\frac{1}{3}(X_i - \mu) + \frac{2}{3}(X_{i+1} - \mu)\right)\left(\frac{1}{3}(X_{i+1} - \mu) + \frac{2}{3}(X_{i+2} - \mu)\right)\right] \\ &= \frac{1}{9}E[(X_i - \mu)(X_{i+1} - \mu)] + \frac{2}{9}E[(X_{i+1} - \mu)^2] \\ &\quad + \frac{2}{9}E[(X_i - \mu)(X_{i+2} - \mu)] + \frac{4}{9}E[(X_{i+1} - \mu)(X_{i+2} - \mu)] \\ &= \frac{2}{9}E[(X_{i+1} - \mu)^2] = \frac{2}{9}\sigma^2 \end{aligned}$$

Εφόσον οι Y_i και Y_{i+1} δεν είναι ασυσχέτιστες, δεν είναι και ανεξάρτητες.

(β) Ναι, είναι όμοια κατανομημένες: Κάθε Y_i είναι το ίδιο σταθμισμένο άθροισμα όμοια κατανομημένων μεταβλητών.

(γ) Χρειάζεται να δείξουμε ότι

$$\lim_{n \rightarrow \infty} P(|M_n - \mu| > \epsilon) = 0$$

για κάθε $\epsilon > 0$. Η μορφή της εξίσωσης είναι παρόμοια με αυτή της ανισότητας του Chebyshev, συνεπώς θα προσπαθήσουμε να τη χρησιμοποιήσουμε στο πρόβλημα.

Η ανισότητα του Chebyshev δίνεται παρακάτω:

$$P(|Y - E(Y)| > a) \leq \frac{\text{var}(Y)}{a^2}.$$

Αν η μέση τιμή της M_n ήταν ίση με μ , η ανισότητα του Chebyshev θα ήταν πολύ χρήσιμη.

$$\begin{aligned} E[M_n] &= \frac{1}{n}E[Y_1 + Y_2 + \cdots + Y_n] \\ &= E[Y_i] = \mu \end{aligned}$$

Υπολογίζουμε τη διακύμανση της M_n ,

$$\begin{aligned} M_n &= \frac{1}{n} \left(\frac{1}{3}X_1 + X_2 + \cdots + X_n + \frac{2}{3}X_{n+1} \right) \\ \text{var}(M_n) &= \frac{1}{n^2} \left(\frac{1}{9}\text{var}(X_1) + \text{var}(X_2) + \cdots + \text{var}(X_n) + \frac{4}{9}\text{var}(X_{n+1}) \right) \\ &= \frac{9n - 4}{9n^2}\sigma^2 \end{aligned}$$

Τελικά, από την ανισότητα του Chebyshev, έχουμε:

$$\begin{aligned}
P(|M_n - \mu| > \epsilon) &\leq \frac{\text{var}(M_n)}{\epsilon^2} \\
P(|M_n - \mu| > \epsilon) &\leq \frac{\frac{9n-4}{9n^2}\sigma^2}{\epsilon^2} \\
\lim_{n \rightarrow \infty} P(|M_n - \mu| > \epsilon) &\leq \lim_{n \rightarrow \infty} \frac{(9n-4)\sigma^2}{9n^2\epsilon^2} \\
&= 0
\end{aligned}$$

Συνεπώς, η M_n συγκλίνει ως προς την πιθανότητα στο μ .

Άσκηση 5.

(α) Αφού $Z = 2(X + Y)(X - Y) = 2(X^2 - Y^2)$, θα έχουμε:

$$E[Z] = 2(E[X^2] - E[Y^2]) = 2(\text{Var}(X) - \text{Var}(Y) + E^2[X] - E^2[Y]) = -40$$

(β) Έχουμε ότι:

$$\begin{aligned}
\text{Cov}(T, U) &= \text{Cov}(2X + Y, 2X - Y) \\
&= 4\text{Cov}(X, X) - 2\text{Cov}(X, Y) + 2\text{Cov}(Y, X) - \text{Cov}(Y, Y) \\
&= 4\text{Var}(X) - \text{Var}(Y) = 7
\end{aligned}$$

(γ) Είναι:

$$\begin{aligned}
E[W] &= E[3X + Y + 2] = 3E[X] + E[Y] + 2 = 9 \\
\text{Var}(W) &= \text{Var}(3X + Y + 2) = 9\text{Var}(X) + \text{Var}(Y) + 6\text{Cov}(X, Y) = 48.6 \\
\text{όπου } \text{Cov}(X, Y) &= \rho\sqrt{\text{Var}(X)\text{Var}(Y)} = 3.6.
\end{aligned}$$

Άσκηση 6.

Το σημαντικό σημείο είναι να παρατηρήσουμε ότι οι $X_1, X_2, \dots$ είναι τυχαίες ανεξάρτητες μεταβλητές, και επομένως οι Y_n και Y_{n+k} είναι επίσης ανεξάρτητες αν $k \geq 3$. Άρα, $\text{Cov}(Y_n, Y_{n+k}) = 0$ αν $k \geq 3$. Τώρα για $k = 0$ έχουμε ότι:

$$\text{Cov}(Y_n, Y_n) = \text{Var}(Y_n) = \text{Var}(X_n + X_{n+1} + X_{n+2}) = \text{Var}(X_n) + \text{Var}(X_{n+1}) + \text{Var}(X_{n+2}) = 3\sigma^2$$

όπου έχουμε χρησιμοποιήσει την ανεξαρτησία των X_n, X_{n+1}, X_{n+2} ακόμη μια φορά. Για $k = 1$ έχουμε:

$$\begin{aligned}
\text{Cov}(Y_n, Y_{n+1}) &= \text{Cov}(X_n + X_{n+1} + X_{n+2}, X_{n+1} + X_{n+2} + X_{n+3}) \\
&= \text{Cov}(X_n, X_{n+1} + X_{n+2} + X_{n+3}) + \text{Cov}(X_{n+1}, X_{n+1} + X_{n+2} + X_{n+3}) \\
&= 0 + \text{Cov}(X_{n+1}, X_{n+1} + X_{n+2}) + \text{Cov}(X_{n+1}, X_{n+2}, X_{n+3}) \\
&= \text{Var}(X_{n+1} + X_{n+2}) + 0 = 2\sigma^2
\end{aligned}$$

όπου έχουμε χρησιμοποιήσει τη διγραμμικότητα της συνάρτησης γινομένου απόκλισης και την ανεξαρτησία των X_n, X_{n+1}, X_{n+2} . Τελικά, για $k = 2$ έχουμε:

$$\begin{aligned}
\text{Cov}(Y_n, Y_{n+2}) &= \text{Cov}(X_n + X_{n+1} + X_{n+2}, X_{n+2} + X_{n+3} + X_{n+4}) \\
&= \text{Cov}(X_n + X_{n+1}, X_{n+2} + X_{n+3} + X_{n+4}) + \text{Cov}(X_{n+2}, X_{n+2} + X_{n+3} + X_{n+4}) \\
&= 0 + \text{Cov}(X_{n+2}, X_{n+2}) + \text{Cov}(X_{n+2}, X_{n+3} + X_{n+4}) \\
&= \text{Var}(X_{n+2}) + 0 = \sigma^2
\end{aligned}$$

όπου η λογική είναι παρόμοια με αυτή της περίπτωσης $k = 1$. Περιληπτικά, έχουμε ότι $\text{Cov}(Y_n, Y_{n+k}) = (3 - k)\sigma^2$ για $k = 0, 1, 2$ και $\text{Cov}(Y_n, Y_{n+k}) = 0$ για $k \geq 3$.