ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
Τμήμα Επιστήμης Υπολογιστών

ΗΥ-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες - Εαρινό Εξάμηνο 2009
Διδάσκων: Π. Τσακαλίδης

Λύσεις Πρώτης Σειράς Ασκήσεων

Ημερομηνία Ανάθεσης: 18/02/2009 Ημερομηνία Παράδοσης: 02/03/2009

Άσκηση 1.

(α) Η ροπογεννητρία της τ.μ. $X \sim N(0,\sigma^2)$ είναι $M_X(s) = e^{s^2\sigma^2/2}$. Άρα, για την τ.μ. Y θα ισχύει:

$$M_Y(s) = E(e^{sY}) = E\left(e^{s\sum_{k=0}^{\infty} X_k r^k}\right) = E\left(\prod_{k=0}^{\infty} e^{sX_k r^k}\right) \overset{\text{ανεξάρτητες τ.μ.}}{=} \prod_{k=0}^{\infty} E\left(e^{sX_k r^k}\right)$$

$$\overset{E(e^{s\alpha X})=M_X(s\alpha)}{=} \prod_{k=0}^{\infty} M_X(sr^k) = \prod_{k=0}^{\infty} e^{(sr^k)^2\sigma^2/2} = e^{(s^2\sigma^2/2)\sum_{k=0}^{\infty}(r^2)^k} \overset{r^2\leq 1}{=} e^{s^2\sigma^2/2(1-r^2)}$$

(β) Παρατηρούμε ότι η $M_Y(s)$ είναι η ροπογεννητρία Γκαουσιανής τ.μ. με μέση τιμή 0 και διασπορά $\sigma^2/(1-r^2)$. Η PDF θα μπορούσε επίσης να εκτιμηθεί από το γεγονός ότι η τ.μ. Y είναι ένας γραμμικός συνδυασμός από ανεξάρτητες και όμοια κατανεμημένες Γκαουσιανές τ.μ., και επομένως η Y είναι και αυτή Γκαουσιανή τ.μ. (με μέση τιμή και διασπορά που έχουν ήδη υπολογιστεί).

Άσκηση 2.

(α) Ισχύει ότι:

$$M_V(s) = E(e^{sV}) = E(e^{sXU}) = E_U(E_X(e^{sXU}|U)) \overset{\text{U τ.μ. Bernoulli}}{=}$$

$$= (1-\epsilon)E[e^{sXU}|U=0] + \epsilon E[e^{sXU}|U=1] = (1-\epsilon)E[e^{sX0}] + \epsilon E[e^{sX1}]$$

$$= (1-\epsilon) + \epsilon E[e^{sX}] \overset{\text{X Γκαουσιανή τ.μ.}}{=} (1-\epsilon) + \epsilon\, e^{s^2\sigma^2/2}$$

(β) Χρησιμοποιώντας τους ορισμούς της μέσης τιμής, έχουμε ότι: (ξέρουμε ότι για την τ.μ. U που ακολουθεί Bernoulli κατανομή ισχύει ότι
$E(U^2) = \sum_u u^2 p_U(u) = 0^2 p_U(u=0) + 1^2 p_U(u=1) = 0(1-\epsilon) + 1\epsilon = \epsilon$)

$$E(V) = E(XU) \overset{\text{X, U ανεξάρτητες τ.μ.}}{=} E(X)E(U) = 0E(U) = 0$$

$$E(V^2) = E((XU)^2) = E(X^2U^2) = E(X^2)E(U^2) = \sigma^2\epsilon$$

$$Var(V) = E(V^2) - E(V)^2 = \sigma^2\epsilon$$

Άσκηση 3.

Για να απαντήσουμε τα ερωτήματα (α), (β) και (γ), πρώτα υπολογίζουμε την μέση τιμή και τη διασπορά των Y_n, T_n, και A_n. Άρα, ισχύει ότι:

$$E[Y_n] = E[(1/2)^n X_n] = (1/2)^n E[X] = (1/2)^n 2.$$

$$Var(Y_n) = Var((1/2)^n X_n) \overset{Y=\alpha X+b:\, Var(Y)=\alpha^2 Var(X)}{=} Var(X)(1/2)^{2n} = 9(1/4)^{2n}.$$

$$E[T_n] = E[Y_1 + ... + Y_n] \overset{\text{άθροισμα ανεξάρτητων τ.μ.}}{=} E[Y_1] + ... + E[Y_n] = E[(1/2)X_1] + ... + E[(1/2)^n X_n]$$

$$= (1/2)E[X_1] + ... + (1/2)^n E[X_n] = (1/2)E[X] + ... + (1/2)^n E[X] = (1/2)2 + ... + (1/2)^n 2$$
$$= 2\sum_{i=1}^{n}(1/2)^i = 2(1/2)\frac{1-(1/2)^n}{1-(1/2)} = 2(1-(1/2)^n).$$
$$Var(T_n) = Var(Y_1 + ... + Y_n) = Var((1/2)X_1 + ... + (1/2)^n X_n)$$
$$= (1/2)^2 Var(X_1) + ... + (1/2)^{2n} Var(X_n) = \sum_{i=1}^{n}(1/4)^i Var(X_i) = \sum_{i=1}^{n}(1/4)^i 9$$
$$= 9\sum_{i=1}^{n}(1/4)^i = 9(1/4)\frac{1-(1/4)^n}{1-(1/4)} = 3(1-(1/4)^n).$$
$$E[A_n] = E[(1/n)T_n] = (1/n)E[T_n] = (2/n)(1-(1/2)^n).$$
$$Var[A_n] = Var[(1/n)T_n] = (1/n)^2 Var(T_n) = (3/n^2)(1-(1/4)^n).$$

(α) ΝΑΙ. Η Y_n συγκλίνει στο 0 ως προς την πιθανότητα. Συγκεκριμένα, όσο το n γίνεται πολύ μεγάλο, η μέση τιμή της Y_n προσεγγίζει την τιμή 0 και η διασπορά της Y_n προσεγγίζει επίσης το 0. Άρα, από την ανισότητα του Chebyshev, η Y_n συγκλίνει στο 0 ως προς την πιθανότητα.

(β) ΟΧΙ. Έστω ότι η T_n συγκλίνει ως προς την πιθανότητα σε μία τιμή α. Επίσης, ισχύει ότι:

$$T_n = Y_1 + (Y_2 + ... + Y_n) = Y_1 + ((1/2)^2 X_2 + ... + (1/2)^n X_n) = Y_1 + (1/2)((1/2)X_2 + ... + (1/2)^{n-1} X_n).$$

Παρατηρούμε ότι το $(1/2)X_2 + ... + (1/2)^{n-1}X_n$ συγκλίνει στο ίδιο όριο με την T_n καθώς $n \to \infty$. Εάν η T_n συγκλίνει στο α τότε η Y_1 θα πρέπει να συγκλίνει στο $\alpha/2$. Αλλά αυτό είναι λάθος, οπότε έρχεται σε αντίθεση με την αρχική μας υπόθεση.

(γ) ΝΑΙ. Η A_n συγκλίνει στο 0 ως προς την πιθανότητα. Όσο το n μεγαλώνει, η μέση τιμή της A_n προσεγγίζει το 0 και η διασπορά της A_n επίσης προσεγγίζει το 0. Οπότε, από την ανισότητα του Chebyshev η A_n συγκλίνει στο 0 ως προς την πιθανότητα. (Αυτό θα μπορούσαμε να το αποδείξουμε χρησιμοποιώντας τον Ασθενή Νόμο των Μεγάλων Αριθμών WLLN παρατηρώντας ότι οι τ.μ. A_n είναι i.i.d. με πεπερασμένη μέση τιμή και διασπορά).

Άσκηση 4.

(α) Ισχύει ότι: $A_n = \frac{X_1+...+X_n}{n}$. Η ανισότητα του Chebyshev δίνει το εξής:

$$P(|A_n - E[A_n]| \geq 0.02) \leq \frac{Var(A_n)}{0.02^2},$$

όπου

$$E[A_n] = \frac{E[X_1 + ... + X_n]}{n} = E[X]$$

και

$$Var(A_n) = Var((X_1 + ... + X_n)/n) = (1/n^2)(Var(X_1) + ... + Var(X_n))$$
$$= nVar(X)/n^2 = Var(X)/n = 9/n.$$

Αυτό μας δίνει τον Ασθενή Νόμο των Μεγάλων Αριθμών (WLLN):

$$P(|A_n - E[X]| \geq 0.02) \leq (9/n)/0.02^n = 0.318.$$

Λύνοντας ως πρός n έχουμε ότι $n = 70754.7$ και αφού το n θα πρέπει να είναι ακέραιος $n = 70755$.

(β) Το Κεντρικό Οριακό Θεώρημα (CLT) μας επιτρέπει να προσεγγίσουμε την τ.μ. $S_n = X_1 + ... + X_n$ σαν μία τ.μ. που ακολουθεί κανονική κατανομή. Μας ενδιαφέρει η ανισότητα

$$P(|A_n - E[X]| > 0.02) < 0.318$$

όπου η τ.μ. $Y_n = (A_n - E[X]) = (\frac{X_1+...+X_n}{n} - E[X]) = (S_n - nE[X])/n$ ακολουθεί επίσης κανονική κατανομή με $E[Y_n] = 0$ και $Var(Y_n) = Var(A_n) = 9/n$. Οπότε,

$$P(|Y_n| > 0.02) = P((Y_n > 0.02) \cup (Y_n < -0.02)) = P(Y_n > 0.02) + P(Y_n < -0.02)$$

$$= \Phi(-0.02\sqrt(n)/3) + (1 - \Phi(0.02\sqrt(n)/3)) = 2 - 2(\Phi(0.02\sqrt(n)/3)) \leq 0.318$$

$$= \Phi(0.02\sqrt(n)/3) \geq 0.8410$$

Καθώς η $\Phi(1.00)$ είναι η μικρότερη τιμή από τον πίνακα τιμών της τυπικής κανονικής CDF που είναι μεγαλύτερη από 0.8410 συμπεραίνουμε ότι $n = 22500$.

(γ) Οπότε, το CLT δίνει την προσέγγιση με την μικρότερη δυνατή τιμή του n.

Άσκηση 5.

Η πιθανότητα ότι η πτήση θα είναι "overbooked" ισούται με την πιθανότητα ότι λιγότερα από 20 εισιτήρια σε σύνολο 320 εισιτηρίων θα ακυρωθούν. Έστω N ο αριθμός των εισιτηρίων που ακυρώνονται, ο οποίος είναι διωνυμική κατανομή με παραμέτρους $n = 320$ και $p = 0.1$, έτσι ώστε να ισχύει ότι $E[N]$=320 0.1=32 και $\sigma_N = \sqrt{(320)(0.1)(0.9)} = \sqrt{28.8}$. Χρησιμοποιώντας την προσέγγιση Moivre-Laplace της διωνυμικής, έχουμε ότι

$$P(N < 20) = P(N \leq 19.5) \approx \Phi(-12.5/\sqrt{28.8}) = 1 - \Phi(2.3292) \approx 0.0099.$$

Άσκηση 6.

(α) Στο ερώτημα αυτό χρησιμοποιούμε το Κεντρικό Οριακό Θεώρημα (CLT). Έστω $X_1, ..., X_{400}$, το μήκος του κάθε τμήματος του σωλήνα, ότι είναι i.i.d. τ.μ. με μέση τιμή $1/\lambda = 1/2$ και διασπορά $1/\lambda^2 = 1/4$. Αν $S_{400} = X_1 + ... + X_{400}$ έχουμε ότι:

$$E[S_{400}] = E[X_1 + ... + X_{400}] = 400E[X] = 200$$

$$Var(S_{400}) = Var(X_1 + ... + X_{400}) = 400Var(X) = 100$$

Οπότε, θέλουμε να βρούμε όριο w τέτοιο ώστε $P(S_{400} > w) \approx 0.841$. Από το CLT ισχύει ότι η S_{400} μπορεί να προσεγγιστεί από μία κανονική τ.μ. με μέση τιμή 200 και διασπορά 100.

$$P(S_{400} > w) = P(\frac{S_{400} - 200}{10} > \frac{w - 200}{10}) = \Phi((200 - w)/10) \approx 0.8410$$

Αφού $\Phi(1.00) \approx 0.841$, έχουμε ότι $(200 - w)/10 \approx 1.00$, άρα $w \approx 190$.

(β) Θέλουμε να βρούμε n τέτοιο ώστε $P(S_n > 200) \approx 0.841$, όπου $S_n = X_1 + ... + X_n$. Όπως στο προηγούμενο ερώτημα, έχουμε ότι $E[S_n] = nE[X] = n/2$ και $Var(S_n) = nVar(X) = n/4$. Υπόθέτοντας ότι το n είναι αρκετά μεγάλο έτσι ώστε να μπορούμε να εφαρμόσουμε το CLT, μπορούμε να προσεγγίσουμε ξανά την S_n ως μία κανονική τ.μ. Επομένως έχουμε ότι

$$P(S_n > 200) = P(\frac{S_n - n/2}{\sqrt{n/4}} > \frac{200 - n/2}{\sqrt{n/4}})$$

$$= \Phi(\frac{n/2 - 200}{\sqrt{n/4}}) \approx 0.8410.$$

Αφού $\Phi(1.00) \approx 0.8410$, το n θα είναι περίπου 421.

Άσκηση 7.

Αφού οι τ.μ. Z_j και Θ_j είναι ανεξάρτητες θα ισχύει ότι:

$$E(Z_j \sin\Theta_j) = E(Z_j)E(\sin\Theta_j) = 0,$$

$$\begin{aligned} Var(Z_j \sin\Theta_j) &= E((Z_j \sin\Theta_j)^2) = E(Z_j^2)E((\sin\Theta_j)^2) \\ &= (1/2)(\mu^2 + \sigma^2) \end{aligned}$$

Αφού $Z_j \sin\Theta_j$΄ς είναι i.i.d., από το CLT έχουμε ότι, καθώς $n \to \infty$, θα ισχύει ότι:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{Z_j \sin\Theta_j}{\sqrt{(\mu^2 + \sigma^2)/2}} \sim N(0, 1)$$

Έπειτα, καθώς $n \to \infty$, $X_n \sim N(0, (1/2)(\mu^2 + \sigma^2))$.