

Πανεπιστήμιο Κρήτης - Τμήμα Επιστήμης Υπολογιστών
ΗΥ-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες - Εαρινό Εξάμηνο 2016
Διδάσκων: Π. Τσακαλίδης
Λύσεις 1ης Σειράς Ασκήσεων

Άσκηση 1.

Έχουμε ότι: $W = X + Y$, και $Z = X + 1.2Y$.

Υπολογίζουμε την συνδιασπορά των τυχαίων μεταβλητών W και Z .

$$\text{cov}(W, Z) = \text{cov}(X + Y, X + 1.2Y) = E[(X + Y) \cdot (X + 1.2Y)] - E[X + Y] \cdot E[X + 1.2Y]$$

Υπολογίζουμε τις ποσότητες:

$$E[X + Y] = E[X] + E[Y] = 12$$

$$E[X + 1.2Y] = E[X] + 1.2 \cdot E[Y] = 5 + 1.2 \cdot 7 = 13.4$$

$$\begin{aligned} E[(X + Y) \cdot (X + 1.2Y)] &= E[X^2 + 1.2 \cdot X \cdot Y + X \cdot Y + 1.2Y^2] = \\ &= E[X^2 + 2.2 \cdot XY + 1.2Y^2] = \\ &= E[X^2] + 2.2E[X \cdot Y] + 1.2E[Y^2] = \\ &= 27.4 + 2.2 \cdot E[X \cdot Y] + 1.2 \cdot 51.4 = \\ &= 2.2 \cdot E[X \cdot Y] + 89.08 \end{aligned}$$

Όμως, $\text{var}(X + Y) = 8$. Οπότε έχουμε:

$$\begin{aligned} 8 &= \text{var}(X + Y) \Leftrightarrow \\ 8 &= E[(X + Y)^2] - (E[X + Y])^2 \Leftrightarrow \\ 8 &= E[X^2] + 2 \cdot E[X \cdot Y] + E[Y^2] - (E[X + Y])^2 \Leftrightarrow \\ 8 &= 27.4 + 2 \cdot E[X \cdot Y] + 51.4 - 144 \Leftrightarrow \\ E[X \cdot Y] &= 36.6 \end{aligned}$$

Επομένως, η συνδιασπορά γίνεται:

$$\begin{aligned} \text{cov}(W, Z) &= \text{cov}(X + Y, X + 1.2Y) = \\ &= E[(X + Y) \cdot (X + 1.2Y)] - E[X + Y] \cdot E[X + 1.2Y] = \\ &= 2.2E[X \cdot Y] + 89.08 - 12 \cdot 13.4 = \\ &= 2.2 \cdot 36.6 - 72.72 = 8.8 \end{aligned}$$

Άσκηση 2.

Παρατηρούμε ότι οι τυχαίες μεταβλητές $X_1, X_2, \dots$ είναι ανεξάρτητες. Επομένως, και οι τυχαίες μεταβλητές Y_n και Y_{n+k} θα είναι ανεξάρτητες, για $k \geq 3$.

Επομένως έχουμε: $cov(Y_n, Y_{n+k}) = 0$, για $k \geq 3$

Διακρίνουμε Περιπτώσεις:

1. Για $k = 0 \Leftrightarrow$, χρησιμοποιώντας την ανεξαρτησία των X_n, X_{n+1}, X_{n+2} έχουμε:

$$cov(Y_n, Y_n) = var(Y_n) = var(X_n + X_{n+1} + X_{n+2}) = var(X_n) + var(X_{n+1}) + var(X_{n+2}) = 3\sigma^2$$

2. Για $k = 1 \Leftrightarrow$, χρησιμοποιώντας τη διγραμμικότητα της συνάρτησης γινομένου απόκλισης και την ανεξαρτησία των X_n, X_{n+1}, X_{n+2} έχουμε:

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

3. Για $k = 2 \Leftrightarrow$

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

Τελικά έχουμε,

$$cov(Y_n, Y_{n+k}) = \begin{cases} (3-k)\sigma^2, & k = 0, 1, 2 \\ 0, & k \geq 3 \end{cases}$$

Άσκηση 3.

(α) Αναλύουμε τη ροπογεννήτρια συνάρτηση σε απλά κλάσματα :

$$M_X(s) = \frac{8-3s}{s^2-6s+8} = \frac{8-3s}{(s-4) \cdot (s-2)} = \frac{A}{s-4} + \frac{B}{s-2}$$

Έχουμε ότι:

$$\begin{aligned} 8-3s &= (s-2) \cdot A + (s-4) \cdot B \Leftrightarrow \\ 8-3s &= (A+B) \cdot s + (-2A-4B) \end{aligned}$$

$$\begin{cases} -3 = A+B \\ +8 = -2A-4B \end{cases} \Leftrightarrow A = -2, \quad B = -1$$

Οπότε, η ροπογεννήτρια γίνεται:

$$M_X(s) = \frac{-2}{s-2} + \frac{-1}{s-4} = \frac{1}{2} \left(\frac{4}{4-s} + \frac{2}{2-s} \right)$$

Επομένως, η συνάρτηση πυκνότητας πιθανότητας που προκύπτει από τη ροπογεννήτρια θα είναι:

$$f_X(x) = \frac{1}{2} \cdot (4e^{-4x} + 2e^{-2x})$$

Υπολογίζουμε τη ζητούμενη πιθανότητα:

$$P(X \geq 0.5) = \int_{0.5}^{+\infty} \frac{1}{2} \cdot (4e^{-4x} + 2e^{-2x}) = \frac{e^{-2}}{2} + \frac{e^{-1}}{2}$$

(β) Χρησιμοποιώντας τη ροπογεννήτρια συνάρτηση, η μέση τιμή και η διασπορά της τυχαίας μεταβλητής X γίνονται:

$$E[X] = \frac{d}{ds} M_X(s)|_{s=0} = \frac{d}{ds} \left(\frac{2}{4-s} + \frac{1}{2-s} \right) |_{s=0} = \frac{2}{(4-s)^2} + \frac{1}{(2-s)^2} |_{s=0} = \frac{3}{8}$$

Επίσης, έχουμε:

$$E[X^2] = \frac{d^2}{ds^2} M_X(s)|_{s=0} = \frac{d^2}{ds^2} \left(\frac{2}{4-s} + \frac{1}{2-s} \right) |_{s=0} = \frac{4}{(4-s)^3} + \frac{2}{(2-s)^3} |_{s=0} = \frac{5}{16}$$

Οπότε η διασπορά γίνεται:

$$\text{var}(X) = E[X^2] - (E[X])^2 = \frac{5}{16} - \left(\frac{3}{8}\right)^2 = \frac{11}{64}$$

Άσκηση 4.

(α) Η ροπογεννήτρια συνάρτηση της τ.μ Z γίνεται:

$$\begin{aligned} M_Z(s) &= E[e^{sZ}] = E[e^{s/4(X-3)}] = E[e^{\frac{s}{4} \cdot X} \cdot e^{-3s/4}] = \\ &= e^{-3s/4} \cdot E[e^{\frac{s}{4} \cdot X}] = \\ &= e^{-3s/4} \cdot M_X(s/4) = \\ &= e^{-3s/4} \cdot e^{3s/4 + 8(s/4)^2} = e^{s^2/2} \end{aligned}$$

(β) Η μέση τιμή της τ.μ Z γίνεται:

$$E[Z] = M'_Z(s)|_{s=0} = s \cdot e^{s^2/2} |_{s=0} = 0$$

(γ) Η διασπορά της τ.μ Z γίνεται:

$$\begin{aligned} \text{var}(Z) &= E[Z^2] - (E[Z])^2 = M''_Z(s)|_{s=0} - 0 = (s)' \cdot e^{s^2/2} + s^2 \cdot e^{s^2/2} \cdot \Big|_{s=0} = \\ &= e^{s^2/2} \cdot (1 + s^2) \Big|_{s=0} = 1 \end{aligned}$$

Άσκηση 5.

(α) Χρησιμοποιώντας το Κεντρικό Οριακό Θεώρημα (ΚΟΘ) ορίζουμε:

$X_1, \dots, X_{400}$, το μήκος του κάθε τμήματος του σωλήνα.

Παρατηρούμε ότι είναι ανεξάρτητες τυχαίες μεταβλητές, οι οποίες ακολουθούν την εκθετική κατανομή, με μέση τιμή $\frac{1}{\lambda} = \frac{1}{2}$ και διασπορά: $\frac{1}{\lambda^2} = \frac{1}{4}$. Επίσης, έστω: $S_{400} = X_1 + X_2 + \dots + X_{400}$.

Τότε η μέση τιμή και η διασπορά γίνονται:

$$E[S_{400}] = E[X_1 + X_2 + \dots + X_{400}] = 400 \cdot \frac{1}{2} = 200$$

και:

$$\text{var}[S_{400}] = \text{var}[X_1 + X_2 + \dots + X_{400}] = 400 \cdot \frac{1}{4} = 100$$

Βρίσκουμε ένα όριο για το w τέτοιο ώστε: $P(S_{400} > w) \approx 0.841$.

Χρησιμοποιώντας το Κεντρικό Οριακό Θεώρημα, η S_{400} προσεγγίζεται από μια σχεδόν κανονική κατανομή με μέση τιμή 200 και διασπορά 100.

Οπότε έχουμε:

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

$$\text{Όμως, ισχύει ότι: } \Phi(1.00) \approx 0.841 \Rightarrow \frac{200 - w}{10} \approx 1 \Rightarrow w \approx 190$$

(β) Γνωρίζουμε ότι η πιθανότητα να κατασκευάσουμε μια σωλήνα συνολικού μήκους τουλάχιστον 200, είναι 0.841. Οπότε έχουμε: $P(S_n > 200) \approx 0.841$

Ισχύει ότι: $S_n = X_1 + \dots + X_n$,

με αντίστοιχη μέση τιμή: $E[S_n] = E[X_1] + \dots + E[X_n] = n \cdot \frac{1}{2} = \frac{n}{2}$

και διασπορά: $\text{var}(S_n) = \frac{n}{4}$.

Υποθέτοντας ότι το n είναι αρκετά μεγάλο, εφαρμόζουμε το Κεντρικό Οριακό Θεώρημα, προσεγγίζοντας την S_n ως μια κανονική τ.μ. Επομένως, έχουμε:

$$P(S_n > 200) = P\left(\frac{S_n - n/2}{\sqrt{n/4}} > \frac{200 - n/2}{\sqrt{n/4}}\right) = \Phi\left(\frac{n/2 - 200}{\sqrt{n/4}}\right) \approx 0.841$$

Όμως, ισχύει ότι:

$$\Phi(1.00) \approx 0.841 \Rightarrow \frac{n/2 - 200}{\sqrt{n/4}} = 2 \Rightarrow n \approx 421$$

Άσκηση 6.

Γνωρίζουμε ότι ο συνολικός αριθμός των κουτιών που θέλουμε να μεταφέρουμε είναι $n = 49$, με μέση τιμή $\mu = 205$ και τυπική απόκλιση $\sigma = 15$. Τα 49 κουτιά θα μεταφερθούν επιτυχώς αν ζυγίζουν συνολικά λιγότερο από 9800 κιλά.

Ορίζουμε X το συνολικό αριθμός των κουτιών που πρέπει να μεταφερθούν.

Οπότε ζητούμε την πιθανότητα: $P(X < 9800)$.

Υπολογίζουμε την ανηγμένη τιμή:

$$z = \frac{9800 - 49 \cdot 205}{\sqrt{49 \cdot 15}} = -2.33$$

Χρησιμοποιώντας το Κεντρικό Οριακό Θεώρημα (ΚΟΘ), η ζητούμενη πιθανότητα γίνεται:

$$P(X < 9800) = P(z < -2.33) = 1 - 0.9901 = 0.0099$$

Άσκηση 7.

(α) Η μέση τιμή των τ.μ X_i γίνεται:

$$E[X_i] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5$$

Σύμφωνα με την ανισότητα *Markov* έχουμε:

$$P[S_{100} \geq 400] \leq \frac{100 \cdot 3.5}{400}$$

(β) Υπολογίζουμε τη διασπορά των τ.μ X_i ως εξής:

$$\text{var}(X_i) = E[X_i^2] - (E[X_i])^2$$

Όμως, έχουμε ότι:

$$E[X_i^2] = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = 15$$

Επομένως, η διασπορά γίνεται:

$$\text{var}(X_i) = E[X_i^2] - (E[X_i])^2 = 15 - 3.5^2 = 2.75$$

Σύμφωνα με την ανισότητα του *Chebyshev* έχουμε:

$$P[S_{100} \geq 400] = P[|S_{100} - 350| \geq 50] \leq \frac{100 \cdot 2.75}{50^2} = 0.11$$

(γ) Με βάση το Κεντρικό Οριακό Θεώρημα έχουμε:

$$\begin{aligned} P[S_{100} \geq 400] &= P\left[\frac{S_{100} - 350}{\sqrt{2.75 \cdot 100}} \geq \frac{400 - 350}{\sqrt{2.75 \cdot 100}}\right] \approx \\ &= Q\left(\frac{400 - 350}{\sqrt{2.75 \cdot 100}}\right) \approx \\ &= Q(3.015) \approx \\ &= 1 - \Phi(3.015) \approx \\ &= 0.0009 \end{aligned}$$

Άσκηση 8.

Ο ζητούμενος κώδικας Matlab:

```
%% Initialize variables
N = 1000;
n = 20;
n_bins = 20;
a = 0;
b = 1;

%% 1
X1 = rand(1, N);

%% 2
h1 = figure;
hist(X1, n_bins);

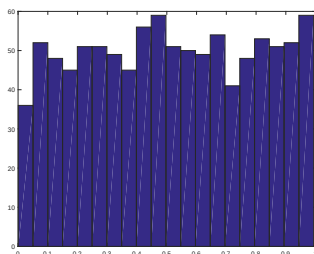
%% 3
X = rand(n, N);

%% 4
Mn = mean(X);
h2 = figure;
hist(Mn, n_bins);

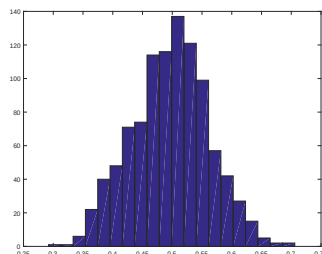
%% 5
mu = (a+b)/2;
sigma = sqrt((1/12) * ((a-b)^2));
Sn = sum(X);
Zn = (Sn - n*mu) / (sigma * sqrt(n));
h3 = figure;
hist(Zn, n_bins);

%% 6
% Run for N=10000, n=100, n_bins=100 changing one variable per experiment
```

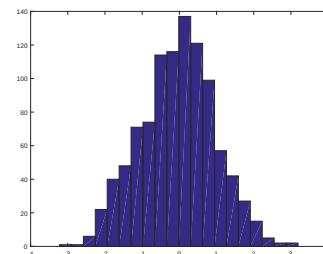
- $N = 1000$, $n = 20$, $nBins = 20$



(α) Η ομοιόμορφη τ.μ. X_1 .



(β) Ιστόγραμμα του δειγματικού μέσου όρου M_n .

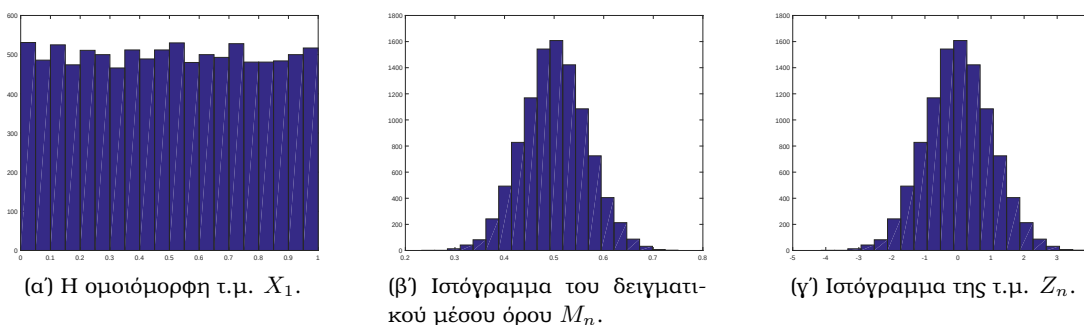


(γ) Ιστόγραμμα της τ.μ. Z_n .

Σχήμα 1

Στην 1α' απεικονίζεται το ιστόγραμμα της ομοιόμορφης τ.μ. X_1 , για $N = 1000$ δείγματα και 20 bins στο διάστημα $[0, 1]$. Στην εικόνα 1β' φαίνεται το ιστόγραμμα του δειγματικού μέσου όρου M_n , για $n = 20$ ανεξάρτητες ομοιόμορφες τ.μ. $X_1, X_2, \dots, X_n$ που μοιάζει πλέον, όπως περιμέναμε από τη θεωρία, να ακολουθεί Γκαουσιανή κατανομή γύρω από τη μέση τιμή $\mu = \frac{a+b}{2} = \frac{0+1}{2} = 0.5$ των X_i . Η εικόνα 1γ' αναπαριστά το ιστόγραμμα του μετασχηματισμού Z_n : Πρώτα έχουμε αφαιρέσει από την S_n , ώστε να προκύψει η τ.μ. $S_n - n\mu$ με μηδενική μέση τιμή και κατόπιν διαιρέσαμε με $\sigma\sqrt{n}$ ώστε η διασπορά να γίνει ίση με τη μονάδα. Τώρα, στο παραγόμενο ιστόγραμμα βλέπουμε στην πράξη το Θεώρημα Κεντρικού Ορίου που μας λέει ότι η ασυμπτωτική μορφή της Z_n τείνει οριακά στην Κανονική κατανομή, $Z_n \sim N(0, 1)$.

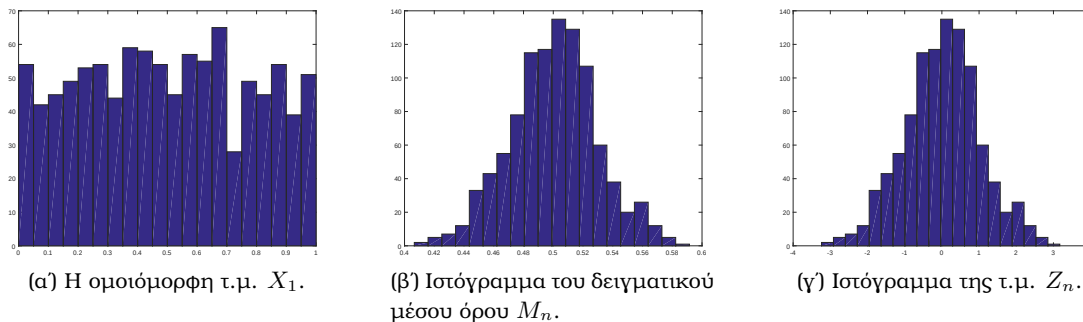
- $N = 10000$, $n = 20$, $nBins = 20$



Σχήμα 2

Αν αυξήσουμε τον αριθμό των δειγμάτων από $N = 1000$ σε $N = 10000$ και αφήσουμε τις υπόλοιπες μεταβλητές ως έχουν, η μόνη διαφορά που παρατηρούμε στα παραγόμενα ιστογράμματα είναι προφανώς μόνο η αύξηση στη συχνότητα εμφάνισης στην εκάστοτε περιοχή τιμών (άξονας y). Δεν παρατηρείται καμία αλλαγή ως προς τη μέση τιμή ή τη διασπορά.

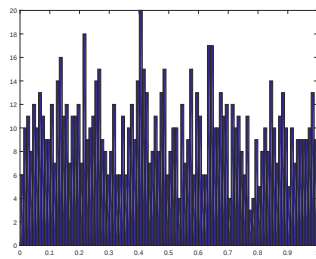
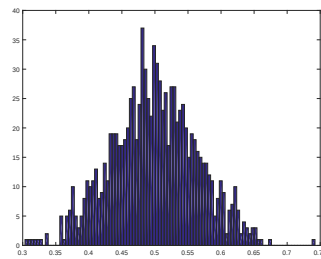
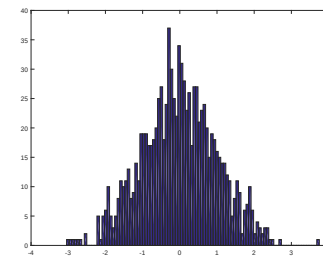
- $N = 1000$, $n = 100$, $nBins = 20$



Σχήμα 3

Αν αυξήσουμε τον αριθμό των μεταβλητών από $n = 20$ σε $n = 100$ και αφήσουμε τις υπόλοιπες μεταβλητές ως έχουν, παρατηρούμε στο παραγόμενο ιστόγραμμα 3β' ότι η διασπορά έχει μειωθεί αισθητά και η κατανομή της M_n έχει τον κύριο όγκο της πλέον πολύ κοντά στη μέση τιμή μ . Επιβεβαιώνεται έτσι η σχετική θεωρία που μας λέει ότι καθώς $n \Rightarrow \infty$, η διασπορά M_n τείνει στο $\emptyset$. Στην εικόνα 3γ', μετά το μετασχηματισμό η διασπορά της Z_n έχει και πάλι επανέλθει στην τιμή 1 προσεγγίζοντας την Κανονική κατανομή.

- $N = 1000$, $n = 20$, $nBins = 100$

(α) Η ομοιόμορφη τ.μ. X_1 .(β) Ιστογράμμο του δειγματικού μέσου M_n .(γ) Ιστογράμμο της τ.μ. Z_n .

Σχήμα 4

Αν αυξήσουμε τον αριθμό των bins από 20 σε 100 και αφήσουμε τις υπόλοιπες μεταβλητές ως έχουν, η μόνη διαφορά που παρατηρούμε στα παραγόμενα ιστογράμματα είναι προφανώς μόνο η μείωση στη συχνότητα εμφάνισης στην εκάστοτε περιοχή τιμών (άξονας y). Τα ιστογράμματα ακολουθούν κατά τα άλλα τις ίδιες κατανομές με αυξημένο αριθμό bins, και δεν παρατηρείται καμία αλλαγή ως προς τη μέση τιμή ή τη διασπορά.