

HY-317: Εφαρμοσμένες Στοχαστικές Διαδικασίες-Εαρινό Εξάμηνο 2021
Διδάσκων: Π. Τσακαλίδης

Λύσεις Δεύτερης Σειράς Ασκήσεων

Άσκηση 1.

(i) Με βάση τη θεωρία, ξέρουμε ότι η τυπική απόκλιση του M_n δίνεται από τον εξής τύπο:

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}$$

Με βάση την εκφώνηση της άσκησης, έχουμε:

$$\sigma_{M_n} \leq 0.01 \Rightarrow \frac{1}{\sqrt{n}} \leq 0.01 \Rightarrow \sqrt{n} \geq \frac{1}{0.01} \Rightarrow \sqrt{n} \geq 100 \Rightarrow n \geq 10000$$

(ii) Με βάση την εκφώνηση της άσκησης, θέλουμε να ισχύει:

$$P(|M_n - h| \leq 0.05) \geq 0.99$$

Κάνοντας χρήση της ανισότητας *Chebyshev*, έχουμε:

$$\begin{aligned} P(|M_n - h| \leq 0.05) &\stackrel{h=E[M_n]}{=} P(|M_n - E[M_n]| \leq 0.05) \\ &= 1 - P(|M_n - E[M_n]| \geq 0.05) \\ &\stackrel{\substack{\sigma_{M_n}^2 = \frac{1}{n} \\ \geq \frac{1}{c=0.05}}}{\geq} 1 - \frac{\frac{1}{n}}{(0.05)^2} \end{aligned}$$

Άρα, τελικά, έχουμε:

$$\begin{aligned} P(|M_n - h| \leq 0.05) \geq 0.99 &\Rightarrow 1 - \frac{\frac{1}{n}}{(0.05)^2} \geq 0.99 \\ &\Rightarrow \frac{\frac{1}{n}}{(0.05)^2} \leq 0.01 \Rightarrow \frac{1}{n} \leq 0.000025 \\ &\Rightarrow n \geq \frac{1}{0.000025} \Rightarrow n \geq 40000 \end{aligned}$$

Άσκηση 2. Εφόσον ζητείται να αποδειχθεί ότι η ακολουθία τ.μ. $Y_1, Y_2, \dots, Y_n$ συγκλίνει κατά πιθανότητα στο 0, θα πρέπει να αποδείξουμε ότι ισχύει η σχέση:

$$\lim_{n \rightarrow +\infty} P(|Y_n - 0| \geq \epsilon) = \lim_{n \rightarrow +\infty} P(|Y_n| \geq \epsilon) = 0 \Rightarrow P(|Y_n| \geq \epsilon) \rightarrow 0$$

(i) Για κάθε $\epsilon > 0$ έχουμε:

$$P(|Y_n - 0| \geq \epsilon) = P(|Y_n| \geq \epsilon) = 0$$

για όλα τα n με $\frac{1}{n} < \epsilon$. Άρα, τελικά είναι: $P(|Y_n| \geq \epsilon) \rightarrow 0$.

(ii) Εφόσον οι $X_1, X_2, \dots, X_n$ είναι ανεξάρτητες τ.μ., έχουμε:

$$\bullet E[Y_n] = E[X_1] \cdot E[X_2] \cdot \dots \cdot E[X_n] = 0 \cdot 0 \cdot \dots \cdot 0 = 0$$

$$\bullet \text{VAR}[Y_n] = E[Y_n^2] - (E[Y_n])^2 = E[Y_n^2] - 0^2 = E[Y_n^2] = E[X_1^2] \cdot E[X_2^2] \cdot \dots \cdot E[X_n^2] = \frac{(1-(-1))^2}{12} \cdot \frac{(1-(-1))^2}{12} \cdot \dots \cdot \frac{(1-(-1))^2}{12} = \frac{4}{12} \cdot \frac{4}{12} \cdot \dots \cdot \frac{4}{12} = \left(\frac{4}{12}\right)^n = \left(\frac{1}{3}\right)^n$$

Από τις παραπάνω σχέσεις, παρατηρούμε ότι ισχύει: $\text{VAR}[Y_n] \rightarrow 0$.

Εφόσον όλες οι Y_n έχουν κοινή μέση τιμή (ίση με 0), κάνοντας χρήση της *ανισότητας Chebyshev*, έχουμε:

$$P(|Y_n - \mu_{Y_n}| \geq c) \leq \frac{\sigma_{Y_n}^2}{c^2} \xrightarrow[\sigma_{Y_n}^2 = (\frac{1}{3})^n]{\mu_{Y_n} = 0} P(|Y_n - 0| \geq c) \leq \frac{(\frac{1}{3})^n}{c^2} \xrightarrow{c=\epsilon} P(|Y_n| \geq \epsilon) \leq \frac{(\frac{1}{3})^n}{\epsilon^2}$$

Άρα, τελικά είναι: $P(|Y_n| \geq \epsilon) \rightarrow 0$.

Άσκηση 3.

(i) Έστω X_i ο χρόνος εκτέλεσης του αλγορίθμου την φορά i , $i = 1, 2, \dots, 400$. Με βάση τα δεδομένα της άσκησης, έχουμε ότι:

- Οι X_i είναι ανεξάρτητες τ.μ.
- $E[X_i] = 17.5$ και $\sigma_{X_i} = 4$

Άρα, για τη ζητούμενη πιθανότητα-με χρήση του *Κεντρικού Οριακού Θεωρήματος*-έχουμε:

$$\begin{aligned} P(S_n < 2 \cdot 60 \cdot 60) &= P(S_n < 7200) \stackrel{n=400}{=} P\left(\frac{S_{400} - 400 \cdot 17.5}{4\sqrt{400}} \leq \frac{7200 - 400 \cdot 17.5}{4\sqrt{400}}\right) \\ &= P\left(\frac{S_{400} - 7000}{80} \leq \frac{7200 - 7000}{80}\right) = P\left(\frac{S_{400} - 7000}{80} \leq \frac{200}{80}\right) \\ &= \Phi(2.5) = 0.9938 \end{aligned}$$

(ii) Δεδομένου ότι ο Γιάννης δεν είναι προετοιμασμένος για το διαγώνισμα της προόδου του HY-317 και απαντά κάθε ερώτηση στην τύχη, η πιθανότητα να επιλέξει την σωστή απάντηση είναι $\frac{1}{10}$ (αφού κάθε multiple-choice ερώτηση έχει 10 δυνατές απαντήσεις).

Ορίζουμε τις τ.μ. X_i ($i = 1, 2, \dots, 50$) ως εξής:

$$X_i = \begin{cases} 3 & \text{με πιθανότητα } p = \frac{1}{10} \\ -1 & \text{με πιθανότητα } 1 - p = \frac{9}{10} \end{cases}$$

Για τις τ.μ. X_i , έχουμε:

- $E[X_i] = 3 \cdot \frac{1}{10} - 1 \cdot \frac{9}{10} = -\frac{6}{10} = -\frac{3}{5}$
- $\text{VAR}[X_i] = E[X_i^2] - (E[X_i])^2 = 3^2 \cdot \frac{1}{10} + (-1)^2 \cdot \frac{9}{10} - \frac{6}{10} = \frac{12}{10} = \frac{6}{5}$

Άρα, η πιθανότητα να πάρει ο Γιάννης προβιβασμό μέσο όρο βαθμολογίας στην πρόοδο του HY-317, θα είναι:

$$\begin{aligned} P\left(\frac{S_n}{n} > 0.5\right) &\stackrel{n=50}{=} P(S_{50} > 50 \cdot 0.5) = P(S_{50} > 25) \\ &= P\left(\frac{S_{50} - 50 \cdot (-\frac{3}{5})}{50\sqrt{\frac{6}{5}}} > \frac{25 - 50 \cdot (-\frac{3}{5})}{50\sqrt{\frac{6}{5}}}\right) \\ &= P\left(\frac{S_{50} + 30}{50\sqrt{\frac{6}{5}}} > \frac{25 + 30}{50\sqrt{\frac{6}{5}}}\right) = P\left(\frac{S_{50} + 30}{50\sqrt{\frac{6}{5}}} > \frac{55}{50\sqrt{\frac{6}{5}}}\right) \\ &= P\left(\frac{S_{50} + 30}{50\sqrt{\frac{6}{5}}} > \frac{11}{10\sqrt{\frac{6}{5}}}\right) = 1 - P\left(\frac{S_{50}}{50\sqrt{\frac{6}{5}}} \leq \frac{11}{10\sqrt{\frac{6}{5}}}\right) \\ &= 1 - \Phi\left(\frac{11}{10\sqrt{\frac{6}{5}}}\right) = 1 - \Phi(1.0042) = 1 - 0.8423 = 0.1577 \end{aligned}$$

Άσκηση 4.

- (i) Έστω $S_n = X_1 + X_2 + \dots + X_n$ ο συνολικός αριθμός αντικειμένων που παράγονται σε n μέρες. Με βάση τα δεδομένα της άσκησης (X_n είναι ανεξάρτητες και όμοια κατανομημένες τ.μ. με μέση τιμή ίση με 5 και διασπορά ίση με 9), έχουμε:

•

$$\begin{aligned} E[S_n] &= E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n] \\ &= 5 + 5 + \dots + 5 = 5n \end{aligned}$$

•

$$\begin{aligned} VAR[S_n] &= VAR[X_1 + X_2 + \dots + X_n] = VAR[X_1] + VAR[X_2] + \dots + VAR[X_n] \\ &= 9 + 9 + \dots + 9 = 9n \end{aligned}$$

•

$$\sigma_{S_n} = \sqrt{VAR[S_n]} = \sqrt{9n} = 3\sqrt{n}$$

Άρα, για τη ζητούμενη προσέγγιση-με χρήση του *Κεντρικού Οριακού Θεωρήματος*-έχουμε:

$$\begin{aligned} P(S_{100} < 440) &= P(S_{100} \leq 439.5) \stackrel{n=100}{=} P\left(\frac{S_{100} - 500}{30} \leq \frac{439.5 - 500}{30}\right) \\ &= P\left(\frac{S_{100} - 500}{30} \leq -2.0167\right) = \Phi(-2.0167) \\ &= 1 - \Phi(2.0167) = 1 - 0.9781 = 0.0219 \end{aligned}$$

- (ii) Εφαρμόζοντας το *Κεντρικό Οριακό Θεώρημα*, έχουμε:

$$\begin{aligned} P(X_1 + X_2 + \dots + X_n \geq 200 + 5n) &\leq 0.05 \Rightarrow P(S_n \geq 200 + 5n) \leq 0.05 \\ \Rightarrow P\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200 + 5n - 5n}{3\sqrt{n}}\right) &\leq 0.05 \Rightarrow P\left(\frac{S_n - 5n}{3\sqrt{n}} \geq \frac{200}{3\sqrt{n}}\right) \leq 0.05 \\ \Rightarrow 1 - P\left(\frac{S_n - 5n}{3\sqrt{n}} \leq \frac{200}{3\sqrt{n}}\right) &\leq 0.05 \Rightarrow 1 - \Phi\left(\frac{200}{3\sqrt{n}}\right) \leq 0.05 \\ \Rightarrow \Phi\left(\frac{200}{3\sqrt{n}}\right) &\geq 0.95 \Rightarrow \Phi\left(\frac{200}{3\sqrt{n}}\right) \stackrel{\text{Normal Tables}}{\geq} \Phi(1.65) \approx 0.95 \\ \Rightarrow \frac{200}{3\sqrt{n}} &\geq 1.65 \Rightarrow \sqrt{n} \leq 40.4040 \\ \Rightarrow n &\leq 1632.483216 \Rightarrow n \leq 1632 \end{aligned}$$

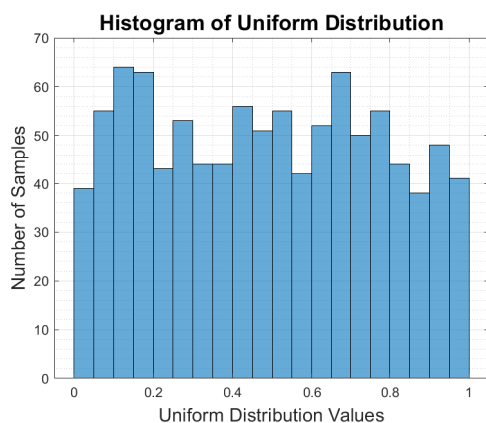
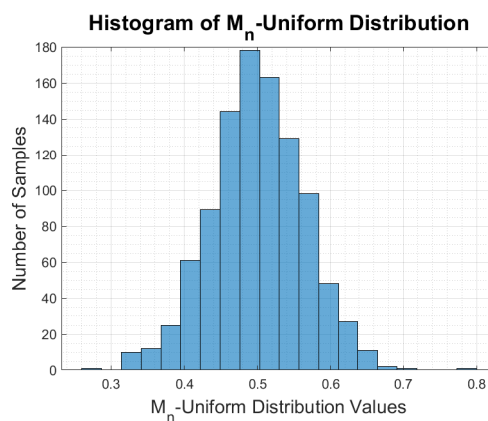
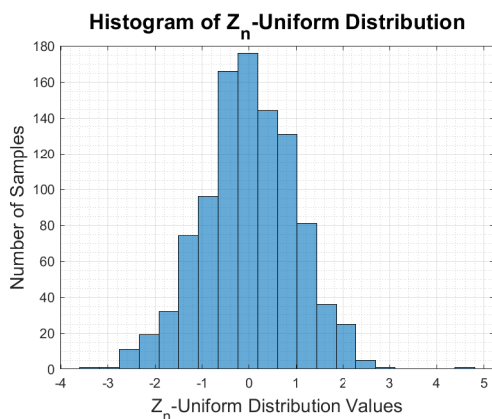
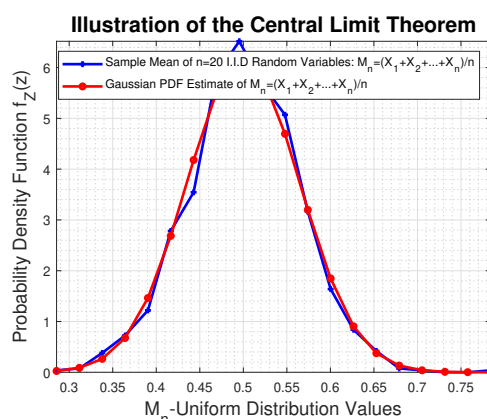
- (iii) Το γεγονός $N \geq 220$ (δηλαδή, ότι χρειάζονται τουλάχιστον 220 ημέρες ώστε ο συνολικός αριθμός παραγόμενων αντικειμένων να υπερβαίνει τα 1000) είναι το ίδιο με το γεγονός $S_{219} \leq 1000$ (δηλαδή, ότι δεν παρήχθησαν περισσότερα από 1000 αντικείμενα τις πρώτες 219 ημέρες).

Έτσι, εφαρμόζοντας το *Κεντρικό Οριακό Θεώρημα*, έχουμε:

$$\begin{aligned} P(N \geq 220) &= P(S_{219} \leq 1000) = P\left(\frac{S_{219} - 5 \cdot 219}{3\sqrt{219}} \leq \frac{1000 - 5 \cdot 219}{3\sqrt{219}}\right) \\ &= P\left(\frac{S_{219} - 1095}{3\sqrt{219}} \leq \frac{1000 - 1095}{3\sqrt{219}}\right) = P\left(\frac{S_{219} - 1095}{3\sqrt{219}} \leq \frac{-95}{3\sqrt{219}}\right) \\ &= P\left(\frac{S_{219} - 1095}{3\sqrt{219}} \leq -2.1398\right) = \Phi(-2.1398) \\ &= 1 - \Phi(2.1398) = 1 - 0.9838 = 0.0162 \end{aligned}$$

Άσκηση 5.

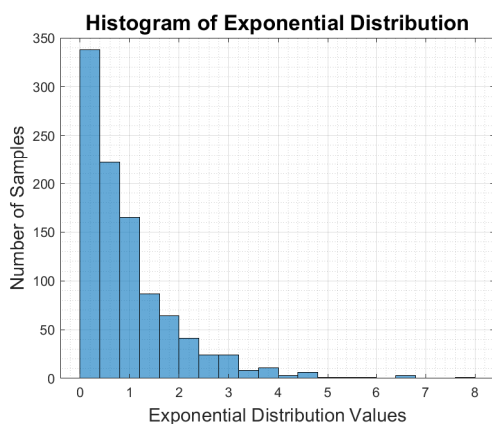
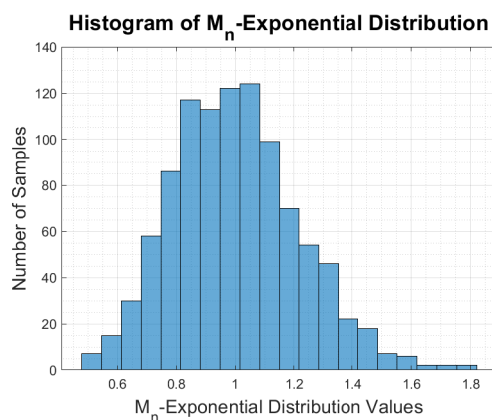
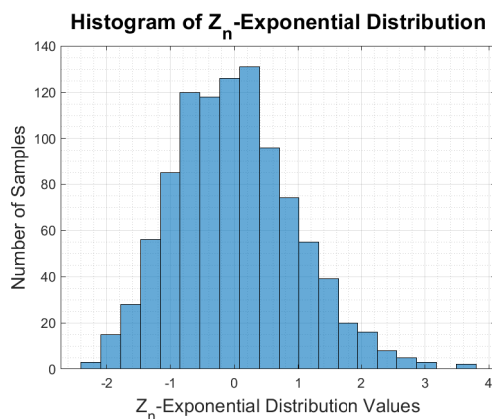
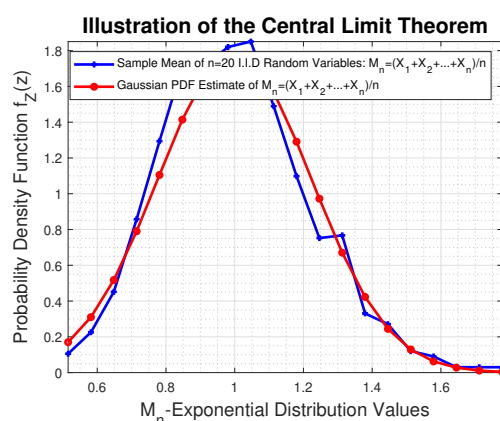
Στο σχήμα 1α' απεικονίζεται το ιστόγραμμα της ομοιόμορφης τ.μ. X_1 , για $N = 1000$ δείγματα και 20 bins στο διάστημα $[0, 1]$. Στο σχήμα 1β' φαίνεται το ιστόγραμμα του δειγματικού μέσου όρου M_n , για $n = 20$ ανεξάρτητες ομοιόμορφες τ.μ. $X_1, X_2, \dots, X_n$ που μοιάζει πλέον, όπως περιμέναμε από τη θεωρία, να ακολουθεί κανονική κατανομή γύρω από τη μέση τιμή $\mu = \frac{a+b}{2} = \frac{0+1}{2} = 0.5$ των X_i . Το σχήμα 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 Κεντρικό Οριακό Θεώρημα, που μας λέει ότι η ασυμπτωτική μορφή της Z_n τείνει οριακά στην Κανονική κατανομή, $Z_n \sim N(0, 1)$. Τέλος, στο σχήμα 1δ' βλέπουμε τις εκτιμήσεις της pdf του δειγματικού μέσου όρου M_n με χρήση ιστογράμματος και κανονικής κατανομής, όπου γίνεται αντιληπτή η καλή προσέγγιση μεταξύ των δύο κατανομών.

(α) Η ομοιόμορφη τ.μ. X .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{X_n} .(γ) Ιστόγραμμα της τ.μ. Z_{X_n} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου M_{X_n} .

Σχήμα 1: Τα ζητούμενα γραφήματα για την ομοιόμορφη Τ.Μ. X για $N = 1000$, $n = 20$, $n_{bins} = 20$.

Ανάλογα συμπεράσματα εξαγονται και για τις εκθετικές τ.μ. $Y_1, Y_2, \dots, Y_n$, μέσω των σχημάτων 2α', 2β', 2γ', 2δ'.

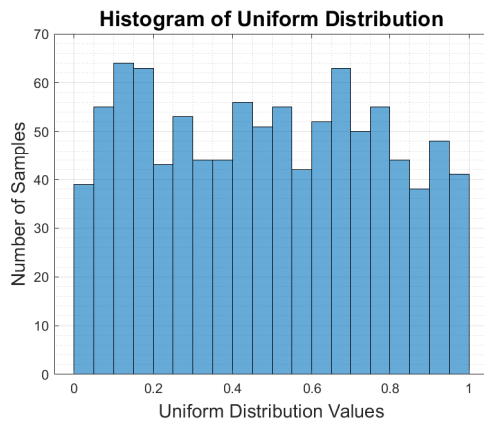
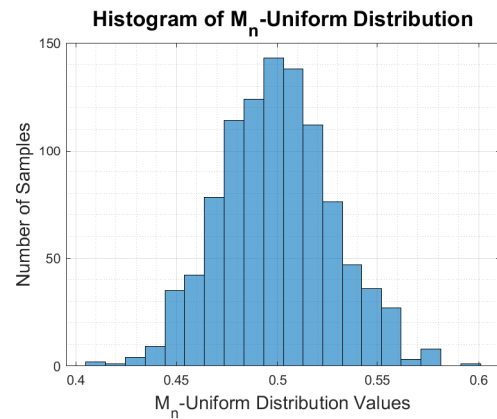
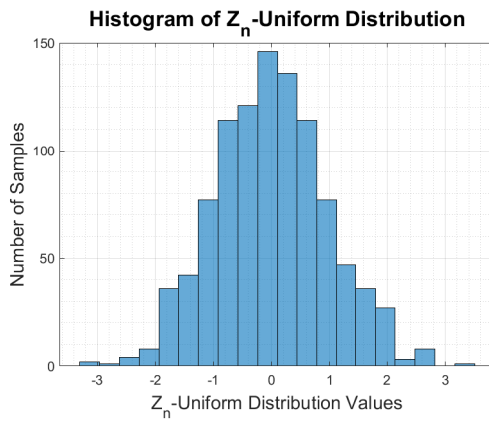
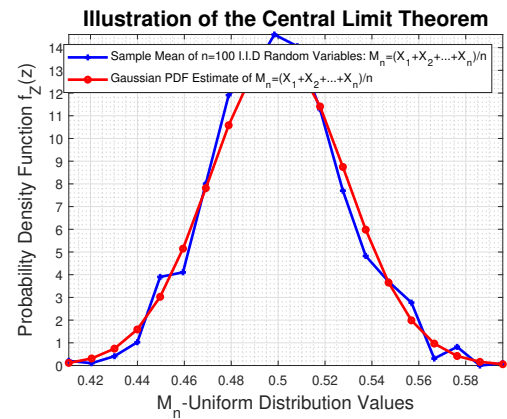
Αν αυξήσουμε τον αριθμό των μεταβλητών από $n = 20$ σε $n = 100$ και αφήσουμε τις υπόλοιπες παραμέτρους ως έχουν, παρατηρούμε στο παραγόμενο ιστόγραμμα 3β' ότι η διασπορά έχει μειωθεί αισθητά και η κατανομή της M_n έχει τον κύριο όγκο της πλέον πολύ κοντά στη μέση τιμή μ . Επιβεβαιώνεται έτσι η σχετική 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. Τέλος, στο σχήμα 3δ' βλέπουμε τις εκτιμήσεις της pdf του δειγματικού μέσου όρου M_n με χρήση ιστογράμματος και κανονικής κατανομής, όπου γίνεται σαφές ότι η διασπορά των κατανομών μειώνεται όσο ο αριθμός των τ.μ. αυξάνει, διατηρώντας παράλληλα την καλή προσέγγιση μεταξύ τους.

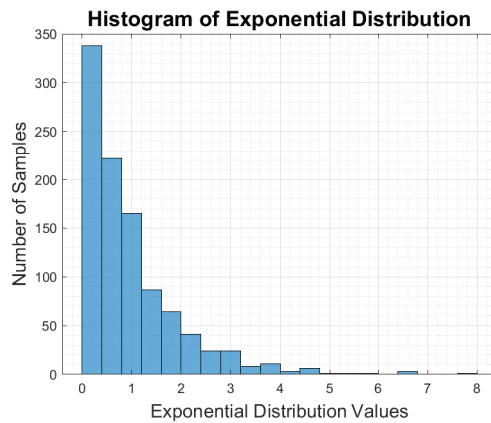
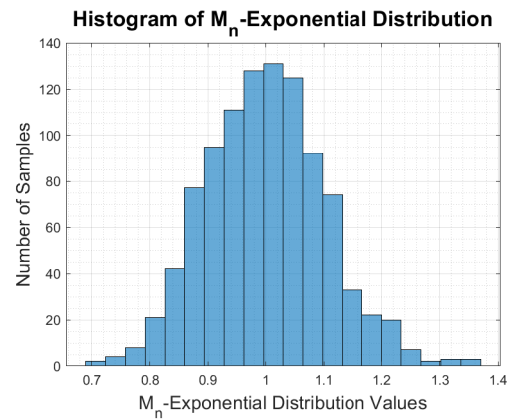
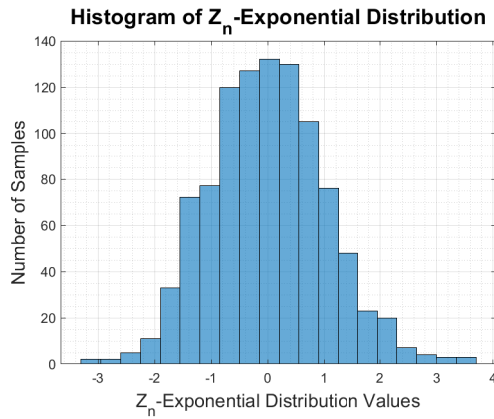
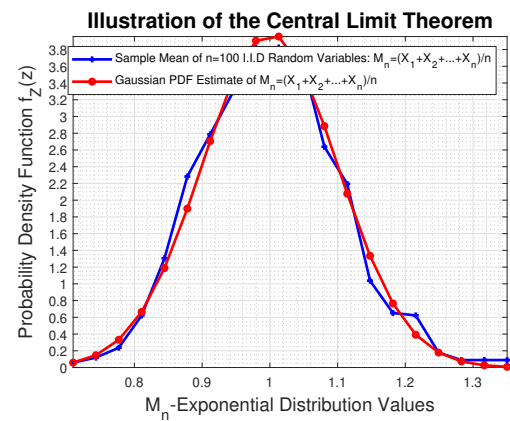
(α') Η εκθετική τ.μ. Y .(β') Ιστογράμμο του δειγματικού μέσου όρου M_{Y_n} .(γ') Ιστογράμμο της τ.μ. Z_{Y_n} .(δ') Εκτίμηση της pdf του δειγματικού μέσου όρου M_{Y_n} .Σχήμα 2: Τα ζητούμενα γραφήματα για την εκθετική Τ.Μ. Y για $N = 1000$, $n = 20$, $n_{bins} = 20$.

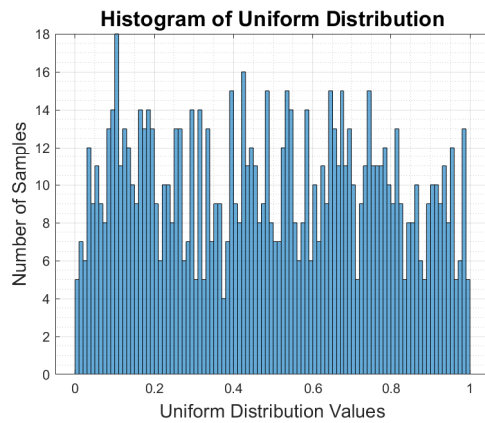
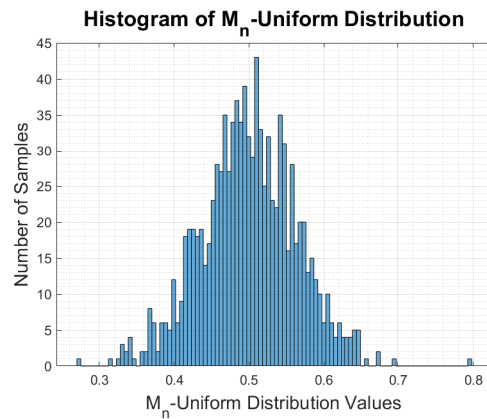
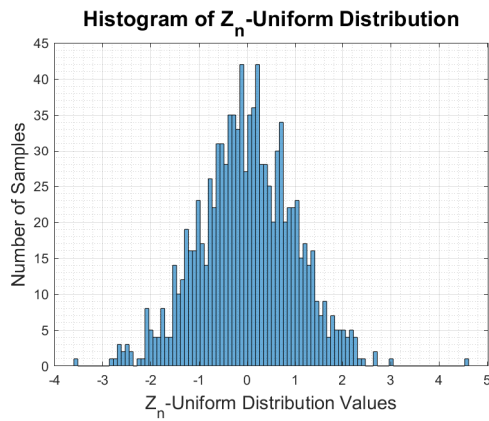
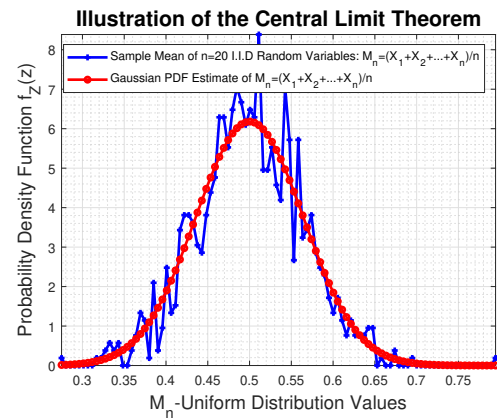
Ανάλογα συμπεράσματα εξαγονται και για τις εκθετικές τ.μ. $Y_1, Y_2, \dots, Y_n$, μέσω των σχημάτων 4α', 4β', 4γ', 4δ'.

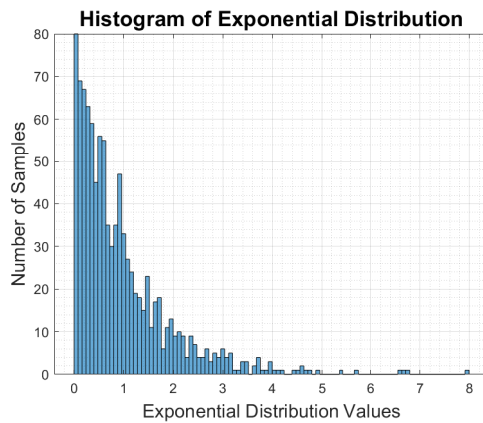
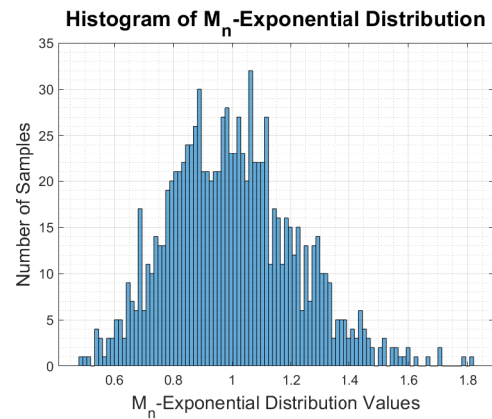
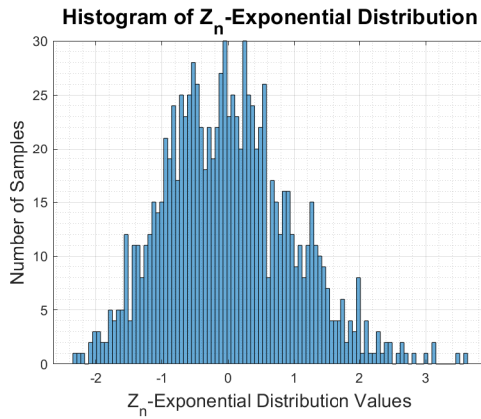
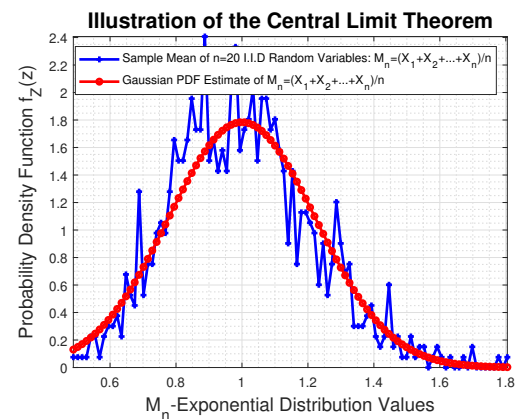
Αν αυξήσουμε τον αριθμό των bins από 20 σε 100 και αφήσουμε τις υπόλοιπες παραμέτρους ως έχουν, η μόνη διαφορά που παρατηρούμε στα παραγόμενα ιστογράμματα είναι προφανώς μόνο η μείωση στη συχνότητα εμφάνισης στην εκάστοτε περιοχή τιμών (άξονας y). Τα ιστογράμματα ακολουθούν κατά τα άλλα τις ίδιες κατανομές με αυξημένο αριθμό bins, και δεν παρατηρείται καμία αλλαγή ως προς τη μέση τιμή ή τη διασπορά. Η αύξηση του αριθμού των bins από 20 σε 100 έχει ως αποτέλεσμα η αντίστοιχη εκτίμηση της pdf που προκύπτει να αποκλίνει περισσότερο από την κανονική. Τα σχετικά γραφήματα φαίνονται στα σχήματα 5 και 6 για την ομοιόμορφη και την εκθετική κατανομή αντίστοιχα.

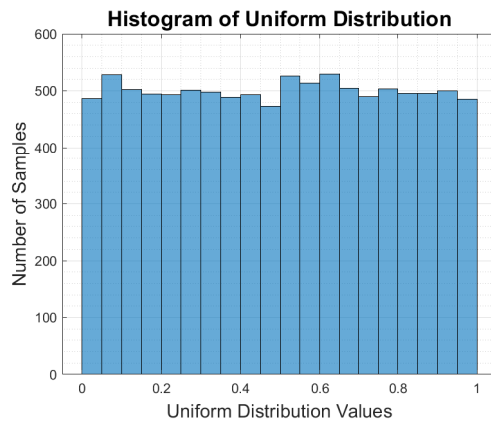
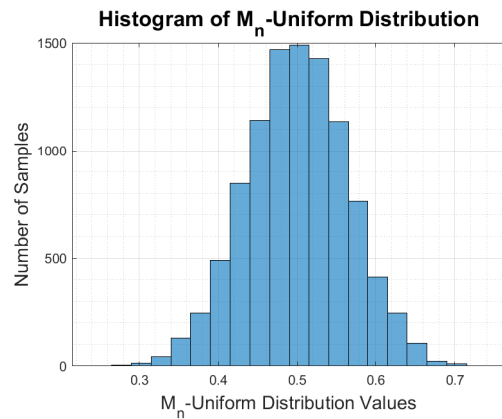
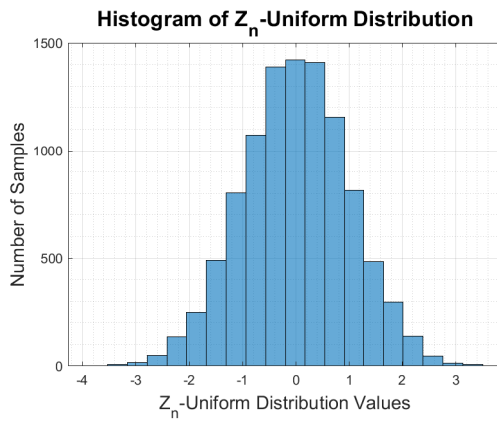
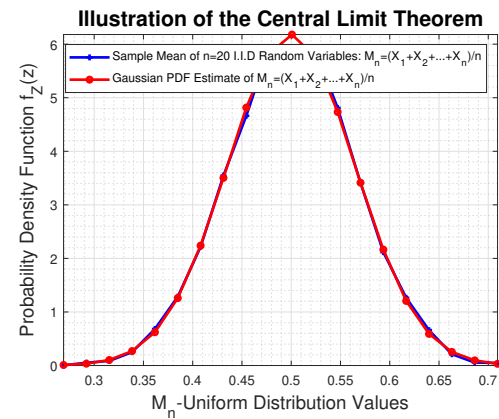
Αν αυξήσουμε τον αριθμό των δειγμάτων από $N = 1000$ σε $N = 10000$ και αφήσουμε τις υπόλοιπες παραμέτρους ως έχουν, η μόνη διαφορά που παρατηρούμε στα παραγόμενα ιστογράμματα είναι προφανώς μόνο η αύξηση στη συχνότητα εμφάνισης στην εκάστοτε περιοχή τιμών (άξονας y). Δεν παρατηρείται καμία αλλαγή ως προς τη μέση τιμή ή τη διασπορά. Η αύξηση του αριθμού των δειγμάτων από $N = 1000$ σε $N = 10000$ έχει ως αποτέλεσμα η αντίστοιχη εκτίμηση της pdf που προκύπτει να συγκλίνει περισσότερο στην κανονική. Τα σχετικά γραφήματα φαίνονται στα σχήματα 7 και 8 για την ομοιόμορφη και την εκθετική κατανομή αντίστοιχα.

(α) Η ομοιόμορφη τ.μ. X .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Xn} .(γ) Ιστόγραμμα της τ.μ. Z_{Xn} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου M_{Xn} .Σχήμα 3: Τα ζητούμενα γραφήματα για την ομοιόμορφη Τ.Μ. X για $N = 1000$, $n = 100$, $n_{bins} = 20$.

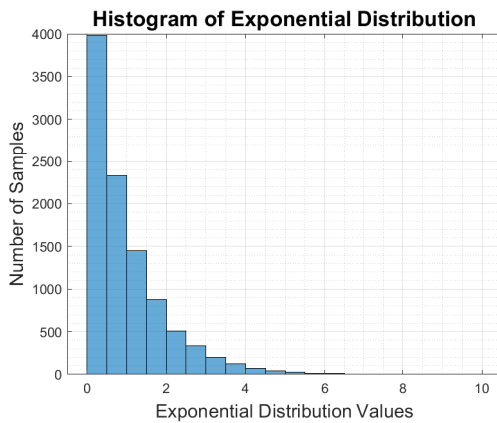
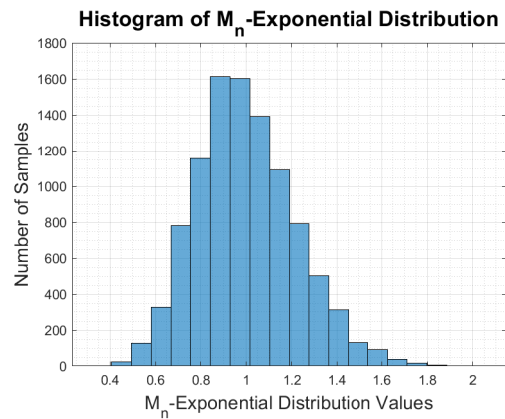
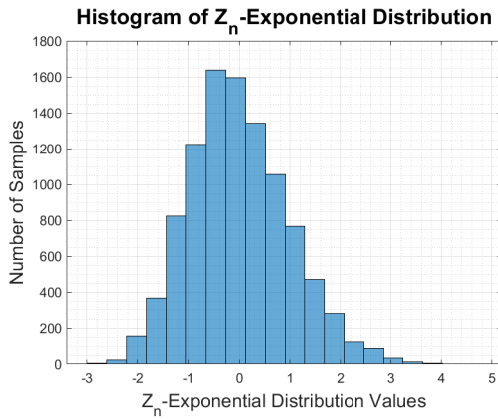
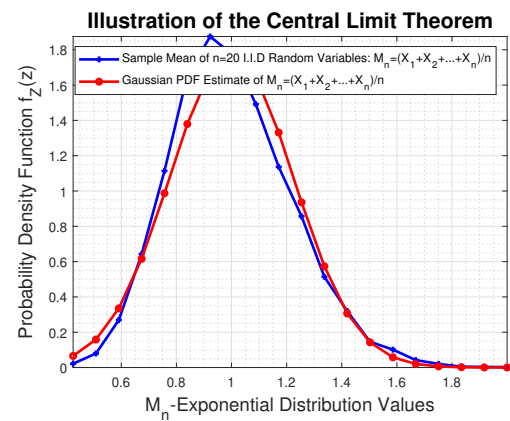
(α) Η εκθετική τ.μ. Y .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Y_n} .(γ) Ιστόγραμμα της τ.μ. Z_{Y_n} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου M_{Y_n} .Σχήμα 4: Τα ζητούμενα γραφήματα για την εκθετική Τ.Μ. Y για $N = 1000$, $n = 100$, $n_{bins} = 20$.

(α) Η ομοιόμορφη τ.μ. X .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Xn} .(γ) Ιστόγραμμα της τ.μ. Z_{Xn} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου M_{Xn} .Σχήμα 5: Τα ζητούμενα γραφήματα για την ομοιόμορφη Τ.Μ. X για $N = 1000$, $n = 20$, $n_{bins} = 20$.

(α) Η εκθετική τ.μ. Y .(β) Ιστογράμμα του δειγματικού μέσου M_{Y_n} .(γ) Ιστογράμμα της τ.μ. Z_{Y_n} .(δ) Εκτίμηση της pdf του δειγματικού μέσου M_{Y_n} .Σχήμα 6: Τα ζητούμενα γραφήματα για την εκθετική Τ.Μ. Y για $N = 1000$, $n = 20$, $n_{bins} = 20$.

(α) Η ομοιόμορφη τ.μ. X .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Xn} .(γ) Ιστόγραμμα της τ.μ. Z_{Xn} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου M_{Xn} .

Σχήμα 7: Τα ζητούμενα γραφήματα για την ομοιόμορφη Τ.Μ. X για $N = 10000$, $n = 20$, $n_{bins} = 20$.

(α) Η εκθετική τ.μ. Y .(β) Ιστόγραμμα του δειγματικού μέσου όρου M_{Y_n} .(γ) Ιστόγραμμα της τ.μ. Z_{Y_n} .(δ) Εκτίμηση της pdf του δειγματικού μέσου όρου M_{Y_n} .Σχήμα 8: Τα ζητούμενα γραφήματα για την εκθετική Τ.Μ. Y για $N = 10000$, $n = 20$, $n_{bins} = 20$.