

Άσκηση 1 (4 μονάδες)

Σε αυτή την άσκηση πρέπει να δημιουργήσετε 4 κατανομές τις οποίες θα τις αποθηκεύσετε σε μορφή πίνακα (matrix). Κάθε κατανομή πρέπει να έχει μεγεθος 100 τιμών, έτσι ώστε ο τελικός σας πίνακας να είναι της μορφής 100 x 4 (τιμές από μία κατανομή σε κάθε στήλη). Ο σκοπός της άσκησης είναι να αποκτήσετε εξοικείωση με τους πίνακες.

Η κατανομές που θα δημιουργήσετε είναι:

- Ομοιόμορφη κατανομή με $\min 0$, $\max 100$
- Κανονική κατανομή με $\text{mean}=50$, $\text{std} = 15$
- Διωνυμική κατανομή με $\text{size}=1000$, $\text{prob}=0.1$
- Poisson κατανομή με $\text{lambda}=10$

Από την στιγμή που θα δημιουργήσετε αυτές κατανομές και θα την τοποθετήσετε στο πίνακα (matrix), όλους υπολογισμούς θα κάνουμε μόνο στο πίνακα αυτό.

Για κάθε κατανομή στο πίνακα υπολογίστε:

1. Μέση τιμή (mean)
2. Τη διαμεσο (median)
3. Μετρήστε το ποσοστό των τιμών που υπάρχουν στο διάστημα: [mean -10, mean+10]
4. Τυπική απόκλιση (sd)
6. Χρησιμοποιήστε την συνάρτηση quantile σε κάθε στήλη ώστε να πάρετε τα quantiles των στηλών. Τα quantiles ορίζονται ως εξής: ένα quantile q που αντιστοιχεί σε a% σημαίνει ότι υπάρχουν a% τιμές μικρότερες από το q. Η διάμεσος σε πιο quantile αντιστοιχεί;

Στο τέλος της άσκησης δημιουργήστε ένα διάνυσμα (vector), οποίος έχει το άθροισμα των τιμών της κάθε γραμμής(row) του πίνακα . Για την λύση αυτή θα χρειαστεί να κάνετε χρήση της επανάληψης (for).

Παρατηρήστε τις παραμέτρους των κατανομών και τις μέσες τιμές που υπολογίσατε ανα στήλη. Υπάρχει κάποια σχέση; (hint: για την κανονική κατανομή, ο μέσος όρος του δείγματος είναι περίπου ίσος με την μέση τιμή της κατανομής).

Επιβεβαιώστε την αντίστοιχη πληροφορία στο wikipedia. Για παράδειγμα για την κανονική κατανομή, η πληροφορία από το wikipedia είναι η εξής:

Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF	$\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right) \right]$
Quantile	$\mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(2p - 1)$
Mean	μ
Median	μ
Mode	μ
...	...

Βλέπετε ότι η μέση τιμή της κατανομής είναι ίση με την παράμετρο μ της κατανομής.

Άσκηση 2 (3 μονάδες)

Κάντε την ίδια διαδικασία με την άσκηση 1, **όμως τώρα στις στήλες του πίνακα βάλτε τιμές από τις αντίστοιχες κατανομές χρησιμοποιώντας την `sample` ως εξής:**

στην πρώτη στήλη 100 τιμές που θα έχετε πάρει από ομοιόμορφη κατανομή

- 20 τιμές από ομοιόμορφη κατανομή με $\min 0$, $\max 100$ (`Unif(0, 100)`)
`sample(20 τιμες από Unif(0, 100), 100)`, δηλαδή δειγματοληψία 100 τιμών από ένα vector 20 τιμών που έχει προκύψει από ομοιόμορφη κατανομή με παραμέτρους 0, 100.
- στην δεύτερη στήλη `sample` 100 τιμών από 20 τιμές της κανονικής της Άσκηση 1
- στην τρίτη στήλη `sample` 100 τιμών από 20 τιμές της διωνυμικής της Άσκησης 1 κοκ.

Υπολογίστε μόνο τους μέσους όρους, τυπικές αποκλίσεις ανά στήλες. Παρατηρείτε διαφορές σε σχέση με τα αντίστοιχα αποτελέσματα της άσκησης 1;

Άσκηση 3 (4)

Θεωρήστε τα παρακάτω μοτίβα

```

AAATCTTACGA
AAATGCTGGGA
CAATCCTACGA
GAATCCTGCGT
CTATAGTGGGA
CCATCCTACGA

```

CAATCCTACGA
CACTTCTACGA
CATTCCTACGA
CAATCCTACGA
GAATCCTACGG
CAACCCGACGA
CAATCCTACGA
CACTCCTACGA
CAATCCTATGA
CAGTCCTACCA
CAGTCCTACGA
CAATGCTACCT
CAATCCTACGG
AAATCCTACCA

Φτιάξτε το PWM που αντιστοιχεί στο μοτίβο αυτό, καθώς και το seqLogo.
Ποιά θέση είναι η περισσότερο και η λιγότερο συντηρημένη;

Σύνολο μονάδων 11/10