Βασικές Έννοιες Στατιστικής

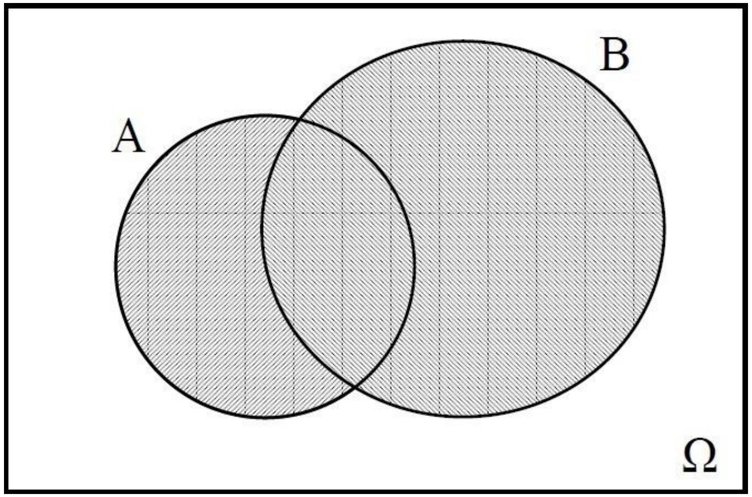
Πιθανότητες - Ορισμοί

- **Πείραμα τύχης** είναι μια διαδικασία που μπορεί να επαναληφθεί όσες φορές θέλουμε, κάτω από τις ίδιες συνθήκες και το αποτέλεσμά του δεν είναι γνωστό εκ των προτέρων. Για παράδειγμα, η γέννηση ενός παιδιού, το τράβηγμα ενός χαρτιού, η ρίψη ενός ζαριού κ.λπ.
- **Δειγματικός χώρος**, Δ, λέγεται το σύνολο των δυνατών αποτελεσμάτων μιας δοκιμής (σημείων του δειγματικού χώρου), π.χ. για τη ρίψη ζαριού Δ={1,2,3,4,5,6}.
- **Γεγονός** Α είναι ένα υποσύνολο του δειγματικού χώρου, π.χ. το υποσύνολο Α={2,4,6} παριστά το γεγονός, το αποτέλεσμα της ρίψης του ζαριού να είναι άρτιος αριθμός.
- Γεγονότα που αποκλείονται αμοιβαία είναι γεγονότα που δεν μπορούν να συμβούν ταυτόχρονα, που δεν έχουν δηλαδή κοινά σημεία στο δειγματικό χώρο: $A_1 \cap A_2 = \varnothing$.

Ένωση $A \cup B$

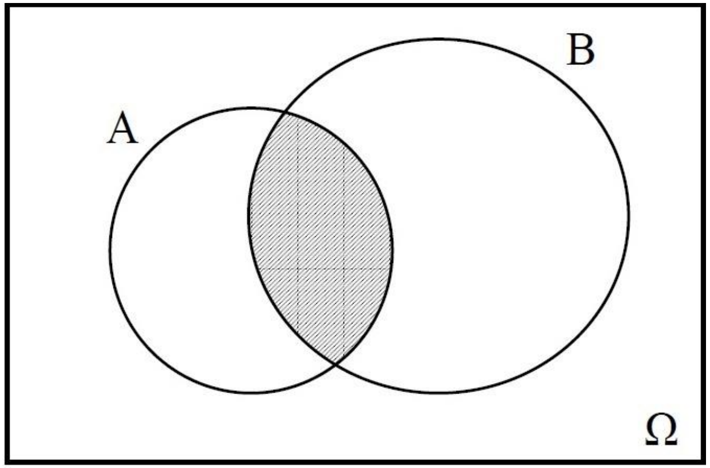
Η ένωση των ενδεχομένων Α και Β περιέχει όλα τα στοιχεία του Ω που ανήκουν είτε στο Α είτε στο Β είτε και στα δύο.

Το $A \cup B$ πραγματοποιείται όταν πραγματοποιηθεί είτε το Α είτε το Β είτε και τα δύο



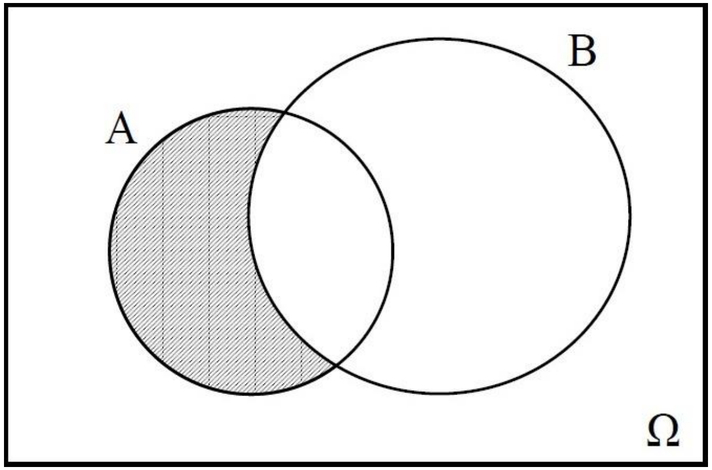
Πιθανότητες - Ορισμοί

Τομή $A \cap B$



Η τομή των ενδεχομένων Α και Β περιέχει όλα τα στοιχεία του Ω που ανήκουν και στο Α και στο Β. Τομή πραγματοποιείται όταν πραγματοποιηθεί και το Α και το Β.

Διαφορά $A\text{-}B$

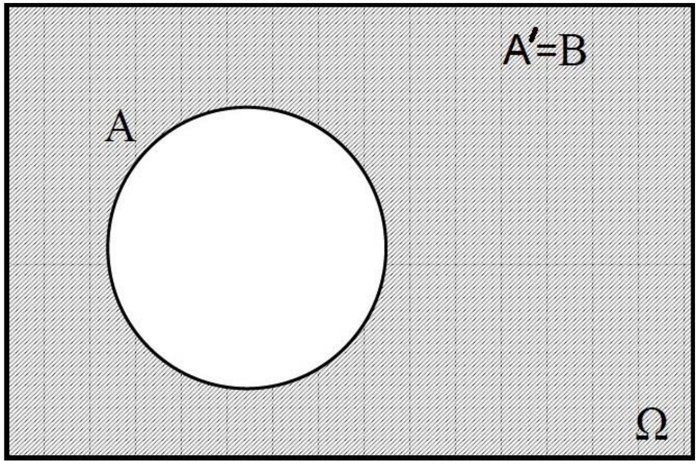


Το ενδεχόμενο που περιέχει όλα τα στοιχεία του *Α* που δεν ανήκουν στο *Β* καλείται διαφορά των *Α* και *Β*. Το *Α-Β* πραγματοποιείται όταν συμβαίνει το Α και όχι το Β.

Πιθανότητες - Ορισμοί

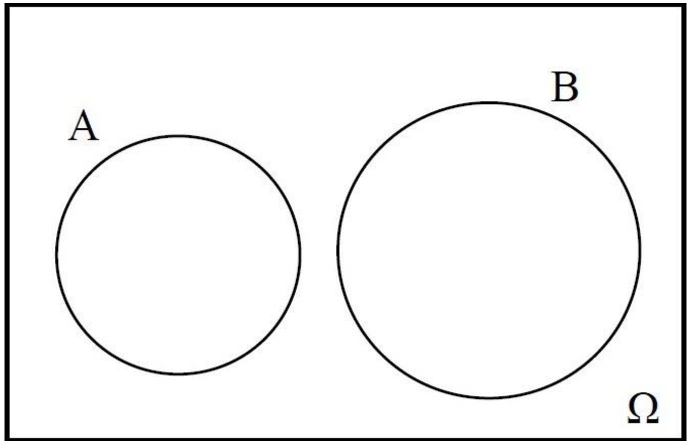
Συμπλήρωμα

Το συμπλήρωμα *A'* ή A^C του *A* ως προς *Ω* περιέχει όλα τα στοιχεία του *Ω* που δεν ανήκουν στο *A.*
Το *A'* πραγματοποιείται όταν δεν πραγματοποιείται το A.



Ξένα ή ασυμβίβαστα ενδεχόμενα

Αν δυο γεγονότα δεν έχουν κανένα κοινό στοιχείο, ονομάζονται ξένα ή ασυμβίβαστα. Η πραγματοποίηση του ενός αποκλείει την πραγματοποίηση του άλλου.



Πιθανότητες - Ορισμοί

Παράδειγμα:
Σε πείραμα τύχης ρίχνουμε δύο φορές ένα νόμισμα. Καλούμε *Α* το ενδεχόμενο «τουλάχιστον μια φορά κεφάλι» και *Β* το ενδεχόμενο «γράμματα στη δεύτερη ρίψη».

Ο χώρος των δειγμάτων είναι *Ω*={*ΚΓ, ΓΚ, ΚΚ, ΓΓ*}
Το ενδεχόμενο *Α* είναι *Α*={*ΚΓ, ΓΚ, ΚΚ*}
Το ενδεχόμενο *Β* είναι *Β*={*ΚΓ, ΓΓ*}

Ισχύει

$$A \cup B = \{ K\Gamma, \Gamma K, KK, \Gamma\Gamma \} = \Omega$$

$$A \cap B = \{ K\Gamma \}$$

$$A' = \{ \Gamma\Gamma \}$$

$$A - B = \{ \Gamma K, KK \}$$

Πιθανότητες - Ορισμοί

Άσκηση
Μετά την ρίψη ενός (τίμιου) ζαριού να προσδιοριστούν τα γεγονότα:
α) A: η ένδειξη να είναι άρτιος αριθμός.
β) B: η ένδειξη να είναι μικρότερη από 3.

Η έννοια της πιθανότητας αναφέρεται σε κάποιο γεγονός. Δηλαδή, μιλάμε για την πιθανότητα πραγματοποίησης κάποιου γεγονότος σε ένα συγκεκριμένο τυχαίο πείραμα. Συμβολίζεται με *P(A)*.

- Η **συνάρτηση πιθανότητας** P(•) ορίζεται ως μια συνάρτηση συνόλου με πεδίο ορισμού (π.ο.) ένα σύνολο γεγονότων Ω και πεδίο τιμών (π.τ.) το [0,1], που ικανοποιεί τα παρακάτω αξιώματα:
    1. $P(A) \geq 0, \ \forall A \in \Omega$.
    2. $P(\Delta)=1$, όπου Δ ο δειγματικός χώρος.
    3. Αν $A_1, A_2, \ldots$ αμοιβαίως αποκλειόμενα γεγονότα του Ω, τότε

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Παράδειγμα: Στη ρίψη ενός αμερόληπτου ζαριού η πιθανότητα καθενός από τα ενδεχόμενα {1},{2},{3},{4},{5} και {6} ισούται με 1/6

Ιδιότητες πιθανοτήτων

- Η πιθανότητα γεγονότος ισούται με το άθροισμα των πιθανοτήτων των σημείων του δειγματικού χώρου που δημιουργούν το γεγονός.
- Το αδύνατο γεγονός συμβολίζεται με $\varnothing$ και έχει μηδενική πιθανότητα, δηλαδή, $P(\varnothing)=0$, π.χ. το γεγονός η ρίψη του ζαριού να δώσει αποτέλεσμα 9 είναι αδύνατο.

- Η πιθανότητα της ένωσης δύο γεγονότων A_1 και A_2, δηλαδή η πιθανότητα να συμβεί το ένα γεγονός ή το άλλο ή και τα δύο, υπολογίζεται από τον **αθροιστικό νόμο των πιθανοτήτων** ο οποίος ισχύει είτε τα γεγονότα είναι αμοιβαίως αποκλειόμενα είτε όχι

$$\mathbf{P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)}$$

- Αν τα γεγονότα A_1 και A_2 είναι **αμοιβαίως αποκλειόμενα**, ισχύει $A_1 \cap A_2 = \varnothing$ και ο αθροιστικός νόμος των πιθανοτήτων γίνεται

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

- Τα γεγονότα A_1, A_2, …, A_m λέγονται **στοχαστικά ανεξάρτητα**, αν και μόνο αν

$$\mathbf{P(A_1 \cap A_2 \cap \ldots \cap A_m) = P(A_1)\, P(A_2) \ldots P(A_m)}$$

Παράδειγμα

Ρίχνουμε δύο “αμερόληπτα” ζάρια. Να βρεθεί η πιθανότητα να φέρουμε ως αποτέλεσμα δύο διαδοχικούς αριθμούς.

Για να βρούμε το δειγματικό χώρο του πειράματος, χρησιμοποιούμε έναν πίνακα “διπλής εισόδου”, όπως φαίνεται στο ακόλουθο σχήμα.

1ο \ 2ο	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Από τον πίνακα αυτόν έχουμε ότι ο δειγματικός χώρος Ω έχει 36 ισοπίθανα δυνατά αποτελέσματα, δηλαδή $N(\Omega)=36$.

•Το ενδεχόμενο A: “να φέρουμε δύο διαδοχικούς αριθμούς”, είναι το $A=\{(1,2),(2,1),(2,3),(3,2),(3,4),(4,3),(4,5),(5,4),(5,6),(6,5)\}$ δηλαδή $N(A)=10$

•Επομένως,

$$P(A)=\frac{N(A)}{N(\Omega)}=\frac{10}{36}=\frac{5}{18} \qquad \frac{5}{18}\approx 0,28$$

Παράδειγμα

Η πιθανότητα ένα οποιοδήποτε προϊόν που παράγεται από μια παραγωγική διαδικασία να είναι ελαττωματικό είναι 0,10. Το τμήμα ποιοτικού ελέγχου εξετάζει τυχαίο δείγμα 3 προϊόντων από αυτή την παραγωγική διαδικασία.

Ποια είναι η πιθανότητα και τα 3 προϊόντα του δείγματος να είναι ελαττωματικά;

- Αφού η πιθανότητα παραγωγής ελαττωματικού είναι σταθερή, η δημιουργία ενός ελαττωματικού δεν επηρεάζεται και δεν επηρεάζει τη δημιουργία άλλων.

- Άρα τα 3 γεγονότα δημιουργίας ελαττωματικών, έστω A_i, i=1,2,3, είναι ανεξάρτητα οπότε η πιθανότητα και τα 3 προϊόντα του δείγματος να είναι ελαττωματικά είναι

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)\, P(A_2)\, P(A_3) = (0,1)\,(0,1)\,(0,1) = 0,001.$$

Πολλές φορές κατά την εκτέλεση ενός πειράματος είναι γνωστή η πραγματοποίηση ενός γεγονότος *B*. Είναι εύλογο το ερώτημα, πόσο επηρεάζει αυτή η πληροφορία την πιθανότητα εμφάνισης ενός άλλου γεγονότος *A*.
Σε αυτές τις περιπτώσεις μιλάμε για τη **δεσμευμένη πιθανότητα** του *A*, όταν γνωρίζουμε ότι το γεγονός *B* έχει πραγματοποιηθεί. Δηλαδή, ζητείται η πιθανότητα πραγματοποίησης του *A* δοθέντος ότι το γεγονός *B* έχει πραγματοποιηθεί. Συμβολίζεται με $P(A|B)$ και ορίζεται:

$$P(A|B)=\frac{P(A\cap B)}{P(B)},\ \ \text{με}\ \ P(B)\neq 0$$

Από τον παραπάνω τύπο προκύπτει εύκολα ο πολλαπλασιαστικός νόμος της πιθανοθεωρίας:

$$P(A\cap B)=P(B)P(A|B)$$

Η πιθανότητα να συμβούν και το *A* και το *B* ισούται με την πιθανότητα να συμβεί το *B* επί την πιθανότητα να συμβεί το *A* υπό την προϋπόθεση να συμβεί το *B*.

Παράδειγμα:
Να βρεθεί η πιθανότητα να έρθει σε μια ρίψη ζαριού αποτέλεσμα μικρότερο του 4, αν α) δε δίνεται άλλη πληροφορία, αν β) είναι γνωστό ότι η ρίψη έδωσε περιττό αριθμό.

α) Έστω *A* το γεγονός «μικρότερο του 4». Επειδή το *A* είναι ένωση των γεγονότων 1, 2, 3, τα οποία είναι ασυμβίβαστα ανά δύο μεταξύ τους έχουμε

$$P(A) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

β) Έστω *B* το γεγονός «περιττός αριθμός». Είναι

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

Ενώ η πιθανότητα να συμβεί και το Α και το Β είναι

$$P(A \cap B) = \frac{2}{6} = \frac{1}{3}$$

Οπότε

$$P\left(A | B\right) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/2} = \frac{2}{3}$$

Συνεπώς δεδομένης της πληροφορίας ότι το αποτέλεσμα είναι περιττό η πιθανότητα αυξάνει.

Πιθανότητες υπό συνθήκη

Θεώρημα Bayes

- Ένα γεγονός μπορεί να έχει διαφορετική πιθανότητα να συμβεί σε κάποιο υποσύνολο του δειγματικού χώρου, από εκείνη που έχει να συμβεί στο σύνολό του. Μια τέτοια πιθανότητα λέγεται πιθανότητα υπό συνθήκη.
- Η πιθανότητα του γεγονότος A_1 δεδομένου του γεγονότος A_2 (υπό τη συνθήκη, δηλαδή ότι ισχύει το A_2), υπολογίζεται από τη σχέση

$$P(A_1|A_2) = P(A_1 \cap A_2) / P(A_2).$$

- Αν τα γεγονότα A_1 και A_2 είναι **στοχαστικά ανεξάρτητα**, τότε προκύπτει ότι

$$P(A_1|A_2) = P(A_1), \text{ αν } P(A_2) > 0.$$

$$P(A_2|A_1) = P(A_2), \text{ αν } P(A_1) > 0.$$

- Έστω $B_1, \ldots, B_k$ αμοιβαίως **αποκλειόμενα γεγονότα** για τα οποία οι πιθανότητες $P(B_i)$, $i=1,2,\ldots,k$ είναι γνωστές και για τα οποία ισχύει $B_1 \cup \ldots \cup B_k = \Delta$, όπου Δ ο δειγματικός χώρος. Έστω επίσης γεγονός A, για το οποίο είναι γνωστές οι υπό συνθήκη πιθανότητες $P(A|B_1), \ldots, P(A|B_k)$. Η πιθανότητα του A υπολογίζεται από τη σχέση

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i).$$

- Οι γνωστές πιθανότητες $P(B_i)$ λέγονται πιθανότητες **εκ των προτέρων** (**a priori**), ενώ οι πιθανότητες $P(B_i|A)$ λέγονται πιθανότητες **εκ των υστέρων** (**a posteriori**) και είναι άγνωστες. Η αναθεώρηση των εκ των προτέρων πιθανοτήτων όταν γίνεται γνωστό ένα επιπλέον στοιχείο (δηλαδή το γεγονός A), πραγματοποιείται με το **θεώρημα του Bayes**

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)}.$$

Παράδειγμα

Οι εταιρείες 1, 2 και 3 προμηθεύουν όργανα μετρήσεων σε ένα εργαστήριο.
Η εταιρεία 1 προμηθεύει το 15% των οργάνων, από τα οποία 2% είναι ελαττωματικά.
Η εταιρεία 2 προμηθεύει το 80%, από τα οποία 1% είναι ελαττωματικά.
Η εταιρεία 3 προμηθεύει το 5%, από τα οποία 3% είναι ελαττωματικά.
Αν κάποιο όργανο βρεθεί ελαττωματικό, ποια η πιθανότητα να προέρχεται από την εταιρεία 3;

- Α ορίζεται το γεγονός ένα τυχαίο όργανο να είναι ελαττωματικό και B_i, i=1,2,3 το γεγονός ένα τυχαίο όργανο να προέρχεται από την εταιρεία i (i=1,2,3).

- Τα B_1, B_2, B_3 είναι αμοιβαίως αποκλειόμενα γεγονότα με εκ των προτέρων πιθανότητες $P(B_1)=0,15$, $P(B_2)=0,80$, $P(B_3)=0,05$ και καλύπτουν το δειγματικό χώρο αφού το άθροισμά τους είναι 1.

- Άρα ισχύει το θεώρημα Bayes και η εκ των υστέρων πιθανότητα που ζητείται $P(B_3|A)$ υπολογίζεται ως εξής

$$P(B_3 \mid A) = \frac{P(B_3 \cap A)}{P(A)} = \frac{P(A \mid B_3)P(B_3)}{\sum_{i=1}^{3} P(A \mid B_i)P(B_i)} = \frac{(0,03)(0,05)}{(0,02)(0,15)+(0,01)(0,80)+(0,03)(0,05)} = 0,12.$$

Τυχαίες μεταβλητές

- Μια μεταβλητή ενός τυχαίου πειράματος ή ελέγχου που μπορεί να λάβει πραγματικές τιμές λέγεται τυχαία μεταβλητή (τ.μ.).
- Οι τ.μ. διακρίνονται σε:
  - **διακριτές** όταν λαμβάνουν ορισμένες συγκεκριμένες τιμές (π.χ. 1, 2, 3, …) και
  - **συνεχείς** όταν λαμβάνουν τιμές από ένα συνεχές υποσύνολο του R (π.χ. R^+).

Παράδειγμα

Κατά τη ρίψη ενός κέρματος τα πιθανά αποτελέσματα είναι κεφάλι (Κ) ή γράμματα (Γ) με ίσες πιθανότητες.

Ποιος ο δειγματικός χώρος που περιλαμβάνει όλα τα αποτελέσματα από τη ρίψη 2 κερμάτων;

Αν Χ τ.μ. ορισμένη ως το πλήθος των Γ σε ρίψη 2 κερμάτων, ορίστε το π.ο. και το π.τ. της;

- Ο δειγματικός χώρος που περιλαμβάνει όλα τα δυνατά αποτελέσματα είναι Δ={ΚΚ,ΚΓ,ΓΚ,ΓΓ}.
- Π.ο. της Χ είναι το Δ.
- Χ(ΚΚ)=0, Χ(ΚΓ)=Χ(ΓΚ)=1, Χ(ΓΓ)=2. Επομένως, το π.τ. R_X της Χ είναι $R_X=\{0,1,2\}$.

Τυχαίες μεταβλητές και κατανομές πιθανότητας

- Για διακριτές τ.μ. X ορίζονται οι:
  - **Συνάρτηση πιθανότητας** ή **κατανομή πιθανότητας** $f(x_k)$. Είναι η συνάρτηση που υπολογίζει την πιθανότητα η τ.μ. X να πάρει μια συγκεκριμένη τιμή x_k:
    $$P(X=x_k) = f(x_k)$$
  - **Αθροιστική συνάρτηση κατανομής** ή **συνάρτηση κατανομής** $F(x_k)$. Είναι η συνάρτηση που υπολογίζει την πιθανότητα η τ.μ. X να πάρει όλες τις τιμές σε ένα διάστημα $(-\infty, x_k]$, όπου x_k ένας οποιοσδήποτε πραγματικός αριθμός:
    $$P(X\leq x_k) = P(-\infty\leq X\leq x_k) = F(x_k)$$
- Για συνεχείς τ.μ. X ορίζονται οι:
  - **Συνάρτηση πυκνότητας πιθανότητας** ή **συνάρτηση πυκνότητας** $f(x)$. Είναι η συνάρτηση με την οποία υπολογίζεται η πιθανότητα η τ.μ. X να πάρει μια τιμή σε ένα διάστημα τιμών $[a, b]$:
    $$P(a \leq X \leq b) = \int_a^b f(x)dx$$
  - **Αθροιστική συνάρτηση κατανομής** ή **συνάρτηση κατανομής** $F(x)$. Είναι η συνάρτηση που υπολογίζει την πιθανότητα η τ.μ. X να πάρει όλες τις τιμές σε ένα διάστημα $(-\infty, x]$, όπου x ένας οποιοσδήποτε πραγματικός αριθμός:
    $$P(X\leq x) = P(-\infty\leq X\leq x) = F(x)$$

Παράδειγμα

Κατά τη ρίψη ενός κέρματος τα πιθανά αποτελέσματα είναι κεφάλι (Κ) ή γράμματα (Γ) με ίσες πιθανότητες.

Ποιος ο δειγματικός χώρος που περιλαμβάνει όλα τα αποτελέσματα από τη ρίψη 3 κερμάτων;

Ποια η πιθανότητα κάθε σημείου;

Αν Χ τ.μ. ορισμένη ως το πλήθος των Κ σε ρίψη 3 κερμάτων, ποια είναι η πιθανότητα να πάρει την τιμή 2, δηλαδή ποια είναι η τιμή της P(X=2);

- Ο δειγματικός χώρος που περιλαμβάνει όλα τα δυνατά αποτελέσματα είναι Δ={ΚΚΚ,ΚΚΓ,ΚΓΚ,ΓΚΚ,ΚΓΓ,ΓΚΓ,ΓΓΚ,ΓΓΓ}.
- Η πιθανότητα κάθε σημείου είναι 1/8.
- Αν Α το γεγονός να έχω 2 Κ κατά τη ρίψη, Α = {ΚΚΓ,ΚΓΚ,ΓΚΚ}, οπότε P(X=2)=P(A)=3/8.

Κατανομές τυχαίων μεταβλητών

Διωνυμική κατανομή

- Σε πολλές περιπτώσεις η διαδικασία ελέγχου αποτελείται από n ανεξάρτητες δοκιμές με 2 δυνατά αποτελέσματα: «επιτυχία» (Ε) ή «αποτυχία» (Α) με σταθερή πιθανότητα επιτυχίας p (και άρα αποτυχίας 1–p) σε κάθε δοκιμή. Μια τέτοια δοκιμή λέγεται δοκιμή (έλεγχος) Bernoulli. Μεταβλητή Bernoulli λέγεται η τ.μ. που παίρνει τιμή 1 με πιθανότητα p (σε Ε) και 0 με πιθανότητα 1–p (σε Α). Η μέση τιμή της κατανομής Bernoulli είναι p.

Συνάρτηση πιθανότητας

$$P(X=c)=\binom{n}{c}p^{c}(1-p)^{n-c},\ c=0,1,...,n$$

Συνάρτηση κατανομής

$$P(X\le c)=\sum_{X=0}^{c}\binom{n}{X}p^{X}(1-p)^{n-X},\ c=0,1,...,n$$

$$\binom{a}{b}=\frac{a!}{b!(a-b)!},\ \text{όπου } a!=1\cdot 2\cdot 3\cdot ...\cdot a$$

$$\text{και } 0!=1,\ \text{εξ' ορισμού}$$

- Η τ.μ. X, που εκφράζει τον αριθμό επιτυχιών σε n ανεξάρτητες δοκιμές Bernoulli, ακολουθεί τη διωνυμική κατανομή.

- Η διωνυμική κατανομή χρησιμοποιείται ευρύτατα για στατιστική ανάλυση αποτελεσμάτων τυχαίας δειγματοληψίας από πρακτικά άπειρους πληθυσμούς με ποσοστό ελαττωματικών p. Τυπική περίπτωση η ανάλυση δειγμάτων από συνεχή παραγωγική διαδικασία που λειτουργεί παράγοντας σταθερό αριθμό ελαττωματικών p, οπότε η πιθανότητα οποιοδήποτε προϊόν της διαδικασίας να είναι ελαττωματικό είναι p.

Παράδειγμα

Σε ένα εργοτάξιο λειτουργούν ανεξάρτητα n=5 ανυψωτικά μηχανήματα.

Κάθε μηχάνημα έχει πιθανότητα βλάβης p=0,20 στη διάρκεια μιας εργάσιμης ημέρας.

Ποια η πιθανότητα να μην υποστούν βλάβη περισσότερα από ένα μηχανήματα στη διάρκεια μιας ημέρας;

- Έχουμε 5 ανεξάρτητα μηχανήματα για καθένα από τα οποία υπάρχει σταθερή πιθανότητα βλάβης p=0,20 (και άρα κανονικής λειτουργίας 0,80). Επομένως η διωνυμική κατανομή μπορεί να μας δώσει τις πιθανότητες βλάβης για το σύνολο των μηχανημάτων.
- Η πιθανότητα να μην υποστεί βλάβη κανένα μηχάνημα υπολογίζεται θεωρώντας n=5, x=0 και p=0,20, και είναι $P(X=0) = 0,328$.
- Η πιθανότητα να υποστεί βλάβη ένα μηχάνημα υπολογίζεται θεωρώντας n=5, x=1 και p=0,20, και είναι $P(X=1) = 0,410$.
- Η πιθανότητα να μην υποστούν βλάβη περισσότερα από ένα μηχανήματα στη διάρκεια της ημέρας υπολογίζεται με χρήση της αθροιστικής συνάρτησης πιθανότητας σύμφωνα με την οποία έχουμε
  $P(X\leq 1) = P(X=0) + P(X=1) = 0,328 + 0,410 = 0,738$.

Κατανομές τυχαίων μεταβλητών

Κανονική κατανομή (1)

- Η κανονική κατανομή γράφεται $N(\mu,\sigma^2)$, όπου μ η μέση τιμή και $\sigma^2>0$ η μεταβλητότητα της τ.μ. Χ που ακολουθεί την κατανομή.

- Η ανηγμένη (τυπική) κανονική μεταβλητή Ζ ορίζεται από το μετασχηματισμό $Z = (X-\mu)/\sigma$ και ακολουθεί την ανηγμένη (τυπική) κανονική κατανομή $N(0,1)$.

- Η συνάρτηση αθροιστικής πιθανότητας $F(x)$ της Χ εξαρτάται από τις παραμέτρους μ και σ^2 και είναι αδύνατο να υπολογιστεί χωρίς τη βοήθεια αριθμητικών μεθόδων. Η πρακτική χρησιμότητα της αναγωγής στη μεταβλητή Ζ συνίσταται στο ότι η συνάρτηση αθροιστικής πιθανότητας $\Phi(z) = P(Z\leq z)$ της Ζ είναι πινακοποιημένη αναλυτικά και επιτρέπει τον υπολογισμό πιθανοτήτων για τη Χ, μετά τον κατάλληλο μετασχηματισμό.

Συνάρτηση πυκνότητας πιθανότητας

- κανονική κατανομή :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \ x \in R$$

- ανηγμένη κανονική κατανομή :

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ z \in R$$

Κατανομές τυχαίων μεταβλητών

Κανονική κατανομή (2)

- Η κανονική κατανομή έχει ιδιαίτερα σημαντικό ρόλο στη στατιστική γενικά και στον έλεγχο ποιότητας ειδικά γιατί προσεγγίζει ικανοποιητικά την κατανομή διαφόρων μεγεθών (π.χ. αποτελέσματα μετρήσεων) σε πολλά πραγματικά φαινόμενα.

- Η κανονική κατανομή έχει την αναπαραγωγική ιδιότητα. Αν X_i ανεξάρτητες κανονικές τ.μ. $N(\mu_i,\sigma_i^2)$, $i=1,\ldots,n$, η $Y=X_1+\ldots+X_n$ είναι επίσης κανονική κατανομή με μέση τιμή $\mu_y=\mu_1+\ldots+\mu_n$ και μεταβλητότητα $\sigma_y^2=\sigma_1^2+\ldots+\sigma_2^2$. Η αναπαραγωγική ιδιότητα γενικεύεται για οποιοδήποτε γραμμικό συνδυασμό ανεξάρτητων κανονικών μεταβλητών.

- Η σπουδαιότητα της κανονικής κατανομής οφείλεται σε μεγάλο βαθμό και στο κεντρικό οριακό θεώρημα σύμφωνα με το οποίο όχι μόνο το άθροισμα ανεξάρτητων κανονικών τ.μ. ακολουθεί κανονική κατανομή, αλλά ακόμα και αν οι ανεξάρτητες τ.μ. ακολουθούν οποιαδήποτε άλλη κατανομή, η κατανομή του αθροίσματός τους τείνει ασυμπτωτικά (όσο δηλαδή $n\rightarrow\infty$) προς την κανονική κατανομή υπό την προϋπόθεση ότι οι μεμονωμένες μεταβλητότητες σ_i^2 είναι μικρές συγκριτικά με το άθροισμά τους, δηλαδή καμιά από τις X_i δεν κυριαρχεί στο άθροισμα Y.

Παράδειγμα

Σε μια αποθήκη ανταλλακτικών αυτοκινήτων ο χρόνος παραμονής των ανταλλακτικών έχει παρατηρηθεί ότι ακολουθεί κανονική κατανομή με μέσο χρόνο παραμονής τους 10 μήνες και διασπορά 9.

Ποια η πιθανότητα να παραμείνει ένα συγκεκριμένο ανταλλακτικό στην αποθήκη μεταξύ 7 και 13 μηνών;

Ποια η πιθανότητα να παραμείνει πάνω από 16 μήνες;

- Ο χρόνος παραμονής των ανταλλακτικών είναι τ.μ. X που ακολουθεί την κανονική κατανομή $N(\mu,\sigma^2)=N(10,9)$. Για λόγους ευκολίας θα χρησιμοποιήσουμε την ανηγμένη κανονική κατανομή $N(0,1)$, οπότε ορίζουμε την ανηγμένη τ.μ. $Z=(X-10)/3$.
- Ζητούμε την πιθανότητα $P(7<X<13)$, οπότε μέσω της ανηγμένης κανονικής κατανομής θα αναζητήσουμε την πιθανότητα
  $P((7-10)/3<Z<(13-10)/3) = P(-1<Z<1)$.
  Από τους σχετικούς πίνακες βρίσκουμε
  $P(-1<Z<1) = P(-1<Z<0) + P(0<Z<1) = 2\ P(0<Z<1) = 2\ (0{,}3413) = 0{,}6826$.
- Ζητούμε επίσης την πιθανότητα $P(Z>16)$, οπότε μέσω της ανηγμένης κανονικής κατανομής θα αναζητήσουμε την πιθανότητα
  $P(Z>(16-10)/3) = P(Z>2)$.
  Από τους σχετικούς πίνακες βρίσκουμε και πάλι
  $P(Z>2) = 0{,}5 - P(0<Z<2) = 1 - [P(Z<0) + P(0<Z<2)] = 0{,}5 - 0{,}4772 = 0{,}0228$.

Στατιστικές εκτιμήσεις

Στην πλειονότητα των περιπτώσεων εξέτασης ενός πληθυσμού με τη βοήθεια τυχαίου δείγματος ενδιαφέρει η εκτίμηση των χαρακτηριστικών τιμών θέσης και διασποράς του πληθυσμού. Συχνά επίσης ενδιαφέρει το ποσοστό ενός πληθυσμού που έχει μια ορισμένη ιδιότητα, π.χ. το ποσοστό ελαττωματικών σε μια παρτίδα παραγωγής.

Οι σημαντικότερες εκτιμήτριες είναι:

- **Εκτιμήτριες θέσης:**
  - Μέση τιμή δείγματος, δηλαδή ο αριθμητικός μέσος όρος των τιμών του δείγματος

$$\boxed{\bar{x} = \sum_{i=1}^{n} x_i \Big/ n}$$

- **Εκτιμήτριες διασποράς:**
  - Τυπική απόκλιση δείγματος
  - Εύρος δείγματος
- **Εκτιμήτρια ποσοστού:**
  - Αν x ο αριθμός των μελών του δείγματος που έχει μια συγκεκριμένη ιδιότητα, η αμερόληπτη εκτιμήτρια του ποσοστού p του πληθυσμού που έχει την ιδιότητα αυτή υπολογίζεται από τη σχέση

$$\boxed{\hat{p} = x/n}$$

Εκτιμήτριες διασποράς

Κύριες εκτιμήτριες της μεταβλητότητας και της τυπικής απόκλισης ενός πληθυσμού είναι αντίστοιχα η μεταβλητότητα και η τυπική απόκλιση του δείγματος. Ενδιαφέρει επίσης και το εύρος του δείγματος:

- Η **μεταβλητότητα δείγματος** S^2 υπολογίζεται από τη σχέση

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

- Η **τυπική απόκλιση δείγματος** S είναι η θετική τετραγωνική ρίζα της μεταβλητότητας δείγματος

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}.$$

- **Εύρος δείγματος** R είναι η διαφορά μεταξύ της μεγαλύτερης (x_{max}) και της μικρότερης (x_{min}) τιμής του δείγματος

$$R = x_{max} - x_{min}.$$

Παράδειγμα

Τυχαίο δείγμα μεγέθους n=12 λαμβάνεται από άγνωστο πληθυσμό.
Οι τιμές του δείγματος είναι 30, 30,5, 31, 31, 32,5, 33, 31,5, 31,5, 30, 31, 31, 29.
Να υπολογιστούν οι κυριότερες εκτιμήτριες θέσης και διασποράς.

- Εκτιμήτριες θέσης
  - Μέση τιμή δείγματος = $(\Sigma x_i)/12 = 372/12 = 31$

- Εκτιμήτριες διασποράς
  - Μεταβλητότητα δείγματος = $(\Sigma(x_i-31)^2)/(12-1) = 13/11 = 1{,}1818$
  - Τυπική απόκλιση δείγματος = $\sqrt{1{,}1818} = 1{,}0871$
  - Εύρος δείγματος = $x_{max}-x_{min} = 33-29 = 4$

Τι μάθαμε;

- Ορισμοί και ιδιότητες πιθανοτήτων.
- Πιθανότητες υπό συνθήκη και Θεώρημα Bayes.
- Τυχαίες μεταβλητές.
- Συνάρτηση πιθανότητας διακριτής τυχαίας μεταβλητής.
- Συνάρτηση πυκνότητας πιθανότητας συνεχούς τυχαίας μεταβλητής.
- Συνάρτηση αθροιστικής πυκνότητας πιθανότητας διακριτής και συνεχούς τυχαίας μεταβλητής.
- Κατανομές τυχαίων μεταβλητών: διωνυμική, κανονική.
- Στατιστικές εκτιμήτριες θέσης τυχαίων μεταβλητών, διασποράς πληθυσμού και ποσοστού πληθυσμού που έχει κάποια συγκεκριμένη ιδιότητα.