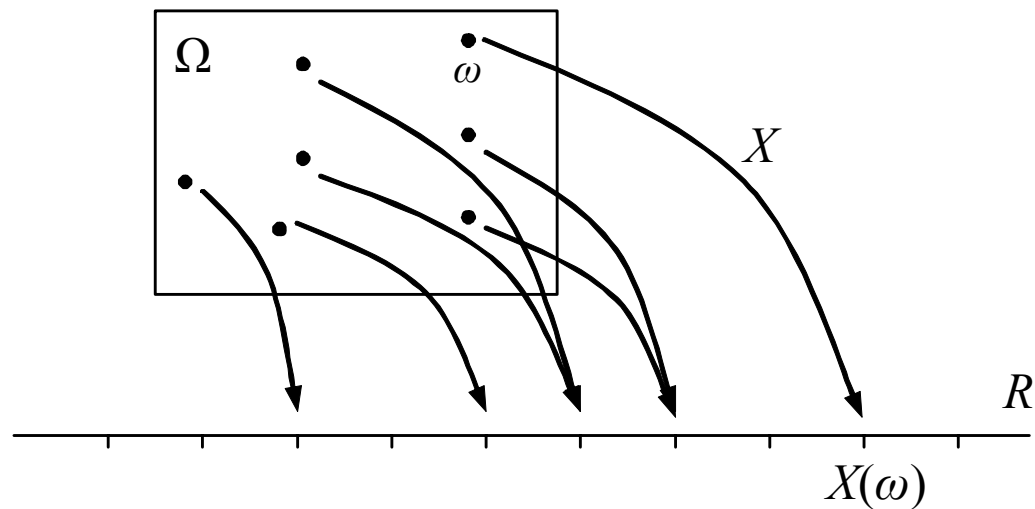

2. Τυχαίες Μεταβλητές.

- Κατά τη μελέτη ενός τυχαίου πειράματος, ενδιαφερόμαστε συνήθως για *κάποια συνάρτηση του αποτελέσματος* και όχι για το αποτέλεσμα αυτό καθαυτό

(π.χ. σε μια ακολουθία δοκιμών, μας ενδιαφέρει το πλήθος από επιτυχίες και όχι ποιες δοκιμές ήταν επιτυχίες).

- Θα μελετήσουμε λοιπόν μεταβλητές των οποίων η τιμή καθορίζεται από το αποτέλεσμα κάποιου στοχαστικού πειράματος. Για το λόγο αυτό οι μεταβλητές αυτές καλούνται *τυχαίες μεταβλητές*.

- Κάθε απεικόνιση X από το δειγματικό χώρο Ω σε κάποιο υποσύνολο των πραγματικών αριθμών R θα καλείται **τυχαία μεταβλητή (τ.μ.)**. (Η τ.μ. X απεικονίζει κάθε στοιχείο ω του Ω στο $X(\omega) \in R$)

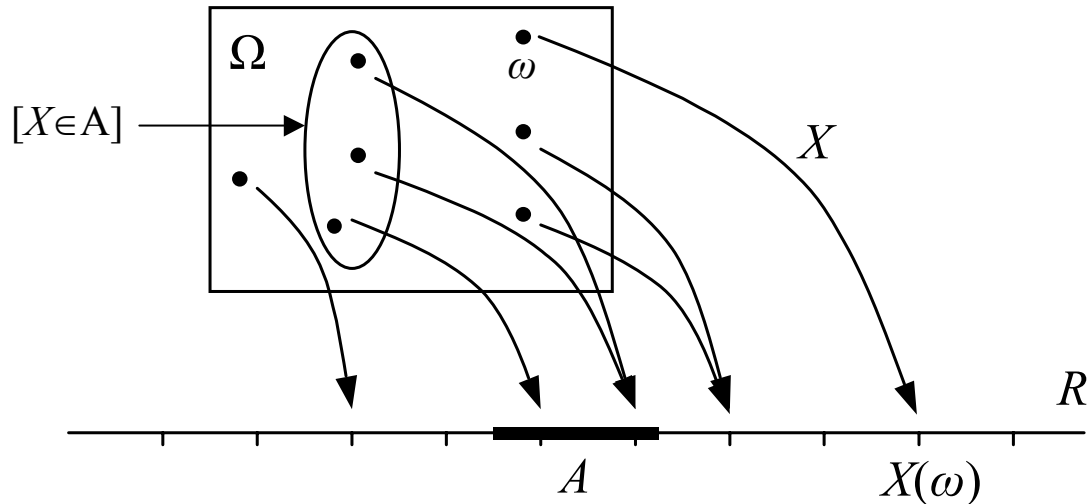


- Γράφουμε $[X \in A]$ για κάποιο $A \subseteq \mathbb{R}$ εννοώντας το ενδεχόμενο

$$[X \in A] = \{\omega \in \Omega: X(\omega) \in A\}.$$

- Γράφουμε $P([X \in A])$ ή απλούστερα $P(X \in A)$ εννοώντας την πιθανότητα

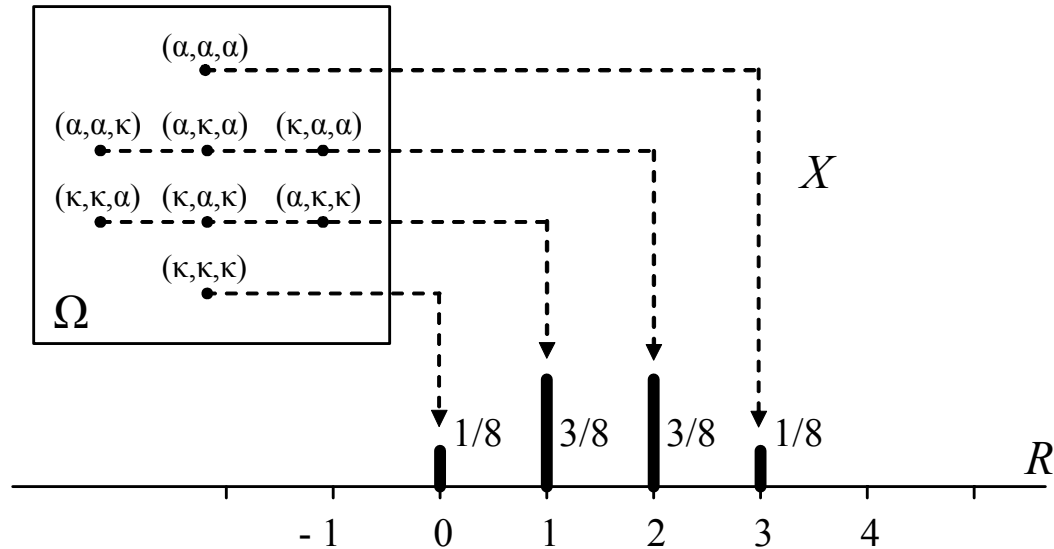
$$P(X \in A) = P(\{\omega \in \Omega: X(\omega) \in A\}).$$



(Για να μην δημιουργούνται τεχνικά προβλήματα, απαιτούμε η απεικόνιση X να είναι τέτοια ώστε τα σύνολα της μορφής $[X \in A]$ να είναι πράγματι ενδεχόμενα του Ω).

Παράδειγμα. Έστω μία οικογένεια με τρία παιδιά. Ας θεωρήσουμε την τ.μ. X η οποία εκφράζει το πλήθος των αγοριών. Ο δειγματικός χώρος θα είναι

$$\Omega = \{(\alpha, \alpha, \alpha), (\alpha, \alpha, \kappa), (\alpha, \kappa, \alpha), (\kappa, \alpha, \alpha), (\alpha, \kappa, \kappa), (\kappa, \alpha, \kappa), (\kappa, \kappa, \alpha), (\kappa, \kappa, \kappa)\}$$



Αν θεωρήσουμε ότι τα 8 στοιχειώδη ενδεχόμενα είναι ισοπίθανα τότε

$$P(X=0) = P(\{\omega \in \Omega: X(\omega)=0\}) = P(\{(\kappa, \kappa, \kappa)\}) = 1/8,$$

$$P(X=1) = P(\{\omega \in \Omega: X(\omega)=1\}) = P(\{(\alpha, \kappa, \kappa), (\kappa, \alpha, \kappa), (\kappa, \kappa, \alpha)\}) = 3/8,$$

$$P(X=2) = P(\{\omega \in \Omega: X(\omega)=2\}) = P(\{(\alpha, \alpha, \kappa), (\alpha, \kappa, \alpha), (\kappa, \alpha, \alpha)\}) = 3/8,$$

$$P(X=3) = P(\{\omega \in \Omega: X(\omega)=3\}) = P(\{(\alpha, \alpha, \alpha)\}) = 1/8.$$

- Παρατηρούμε ότι η συνολική πιθανότητα (1 ή 100%) του Ω «κατανέμεται» στα στοιχεία του συνόλου τιμών της τ.μ. X μέσω της συνάρτησης $f(x) = P(X = x)$.

- Στις εφαρμογές, συνήθως δεν μας ενδιαφέρει από ποια στοιχεία του Ω έχει προέλθει η πιθανότητα που έχει κατανεμηθεί στα διάφορα σημεία του R , αλλά μόνο η «κατανομή» της πιθανότητας στον R .

(Για παράδειγμα, συνήθως δεν μας ενδιαφέρει ότι η πιθανότητα $3/8$ στο σημείο 1 έχει προέλθει από τα στοιχειώδη ενδεχόμενα $\omega_1 = (α,κ,κ)$, $\omega_2 = (κ,α,κ)$, $\omega_3 = (κ,κ,α)$, αλλά μόνο το γεγονός ότι στα σημεία 0,1,2,3 έχει κατανεμηθεί πιθανότητα $1/8$, $3/8$, $3/8$ και $1/8$ αντίστοιχα).

- Επομένως, αρκεί να βρούμε ένα τρόπο να περιγράψουμε την «κατανομή» αυτή της συνολικής πιθανότητας στον R .

Η συνάρτηση $F_X: R \rightarrow [0,1]$ ώστε

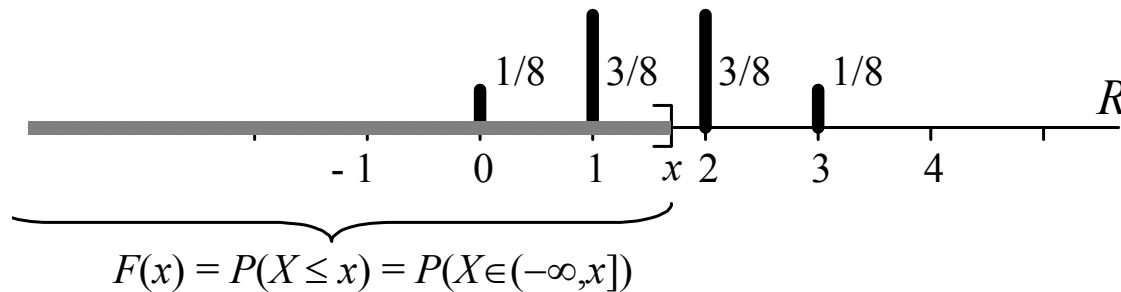
$$F_X(x) = P(X \leq x) = P(X \in (-\infty, x]), \quad x \in R$$

καλείται (αθροιστική) συνάρτηση κατανομής (σ.κ.) της τ.μ. X .

Παράδειγμα Η τ.μ. X εκφράζει το πλήθος των αγοριών σε μία οικογένεια τριών παιδιών. Βρήκαμε παραπάνω ότι

$$P(X=0) = 1/8, \quad P(X=1) = 3/8, \quad P(X=2) = 3/8, \quad P(X=3) = 1/8.$$

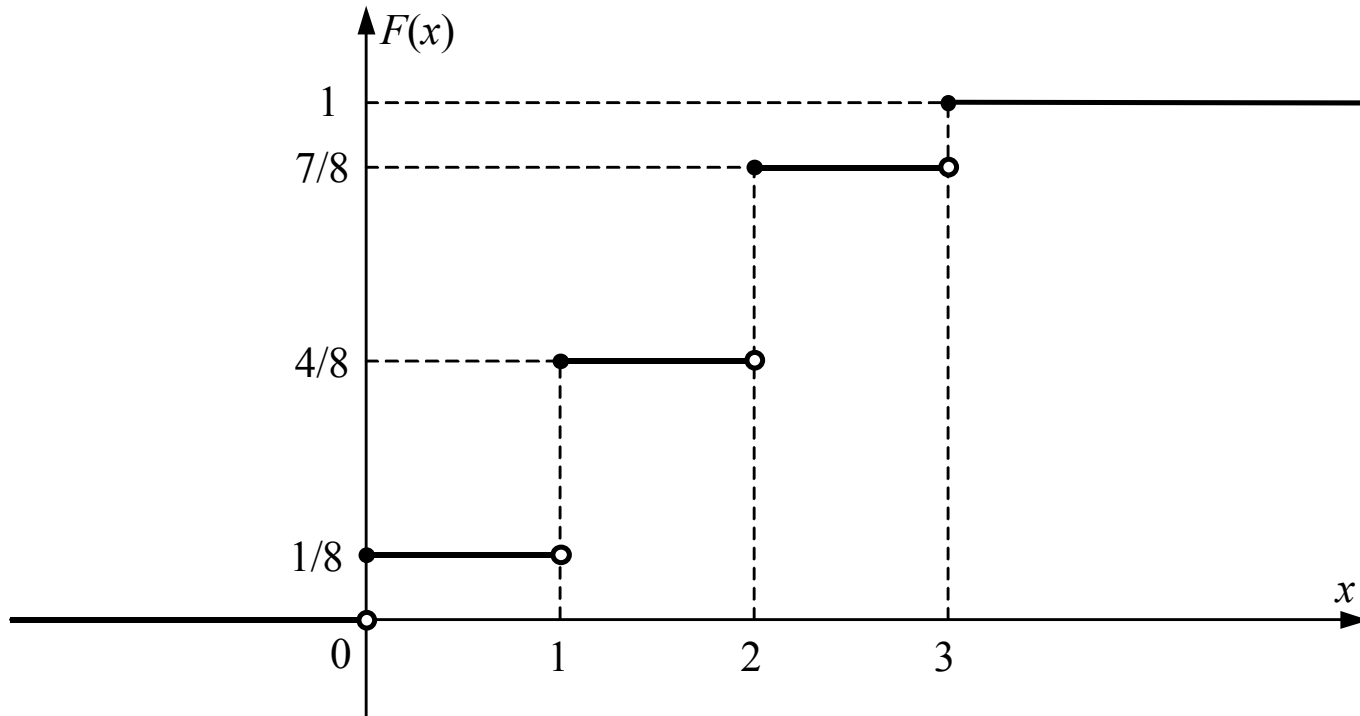
Η $F(x)$ εκφράζει την πιθανότητα που έχει κατανεμηθεί στο σύνολο $(-\infty, x]$.



και συνεπώς θα είναι ίση με

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0, \\ P(X = 0) = 1/8, & 0 \leq x < 1 \\ P(X = 0) + P(X = 1) = 4/8, & 1 \leq x < 2 \\ P(X = 0) + P(X = 1) + P(X = 2) = 7/8, & 2 \leq x < 3 \\ P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 8/8 = 1, & 3 \leq x \end{cases}$$

Σχηματικά θα έχουμε ότι,



- Αποδεικνύεται η επόμενη πρόταση.

Για τη συνάρτηση κατανομής F μιας τ.μ. ισχύουν τα εξής:

(1) $0 \leq F(x) \leq 1$ για κάθε $x \in \mathbb{R}$.

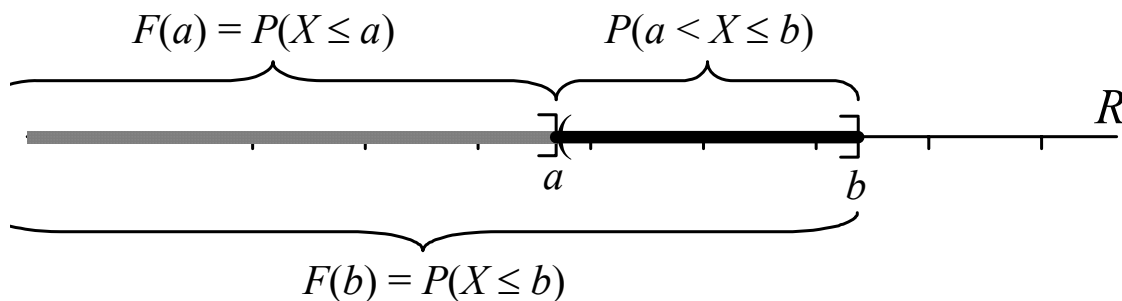
(2) $\lim_{x \rightarrow \infty} F(x) = F(\infty) = 1, \quad \lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0,$

(3) Η F είναι μη φθίνουσα συνάρτηση, δηλαδή για κάθε $x > y$ ισχύει ότι $F(x) \geq F(y)$.

(4) Η F είναι δεξιά συνεχής, δηλαδή $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$.

Επίσης,

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a), \quad a < b$$



- **Διακριτές** καλούνται οι τ.μ. που έχουν ως πεδίο τιμών κάποιο υποσύνολο του $\mathbb{Z}$ ή του $\mathbb{N}$ ή γενικότερα έχουν αριθμήσιμο πεδίο τιμών. Οι αντίστοιχες κατανομές τους θα καλούνται *διακριτές κατανομές*.

(ένα σύνολο καλείται αριθμήσιμο αν μπορεί να γραφεί στη μορφή, $A=\{a_1, a_2, \dots, a_k\}$ (π.χ. $\{1, 2, 3, 5\}$) ή $A=\{a_1, a_2, \dots\}$ (π.χ. $\{0.5, 1, 1.5, 2, \dots\}$ ή $\{3, 6, 9, \dots\}$). Τα σύνολα $\{0, 1, 2, \dots, n\}$, $\mathbb{N}$, $\mathbb{Z}$ είναι αριθμήσιμα ενώ π.χ. ένα διάστημα (α, β) δεν είναι.

- **Συνεχείς** καλούνται οι τ.μ. που έχουν ως πεδίο τιμών ένα διάστημα του $\mathbb{R}$, ή όλο το $\mathbb{R}$ ενώ επιπλέον έχουν παραγωγίσιμες συναρτήσεις κατανομής. Οι αντίστοιχες κατανομές τους θα καλούνται *συνεχείς κατανομές*.

- Υπάρχουν κατανομές που δεν είναι ούτε συνεχείς ούτε διακριτές (π.χ. είναι μικτές ή ιδιάζουσες κατανομές). Τέτοιες κατανομές δεν εμφανίζονται συχνά στις εφαρμογές και δεν θα τις εξετάσουμε στην συνέχεια.

Διακριτές τυχαίες μεταβλητές.

- Στις περισσότερες περιπτώσεις οι διακριτές τ.μ. παίρνουν τιμές στο $\{0,1,2,\dots,n\}$ ή στο $\{0,1,2,\dots\} = \mathbb{N}$.
- Η κατανομή μιας τ.μ. X με τιμές στο A περιγράφεται από την συνάρτηση κατανομής της F . Στην περίπτωση όμως των διακριτών τ.μ. είναι μερικές φορές βολικότερο να χρησιμοποιήσουμε και την λεγόμενη «συνάρτηση πιθανότητας».

Η συνάρτηση $f_X: A \rightarrow [0,1]$,

$$f_X(x) = P(X = x), x \in A$$

καλείται συνάρτηση πιθανότητας (σ.π.) της τ.μ. X .

- Αν $B \subseteq A = X(\Omega)$ τότε,

$$P(X \in B) = \sum_{x \in B} P(X = x) = \sum_{x \in B} f_X(x).$$

- Αν $B = A = X(\Omega)$ τότε προφανώς,

$$\sum_{x \in A} f_X(x) = P(X \in A) = 1.$$

Δηλαδή, το άθροισμα των τιμών της f_X για όλα τα x είναι ίσο με 1. Επίσης, είναι $f_X > 0$ σε ένα αριθμήσιμο υποσύνολο του $\mathbb{R}$.

- Αντίστροφα τώρα, αν μία μη-αρνητική συνάρτηση f παίρνει γνήσια θετικές τιμές σε ένα αριθμήσιμο υποσύνολο του $\mathbb{R}$ και το άθροισμα των τιμών της είναι 1, τότε μπορεί να θεωρηθεί ως η συνάρτηση πιθανότητας μιας (διακριτής) τ.μ.

Σχέση μεταξύ συνάρτησης πιθανότητας και συνάρτησης κατανομής.

- Αν γνωρίζουμε την σ.π. τότε μπορούμε να υπολογίσουμε τη σ.κ. και αντίστροφα.

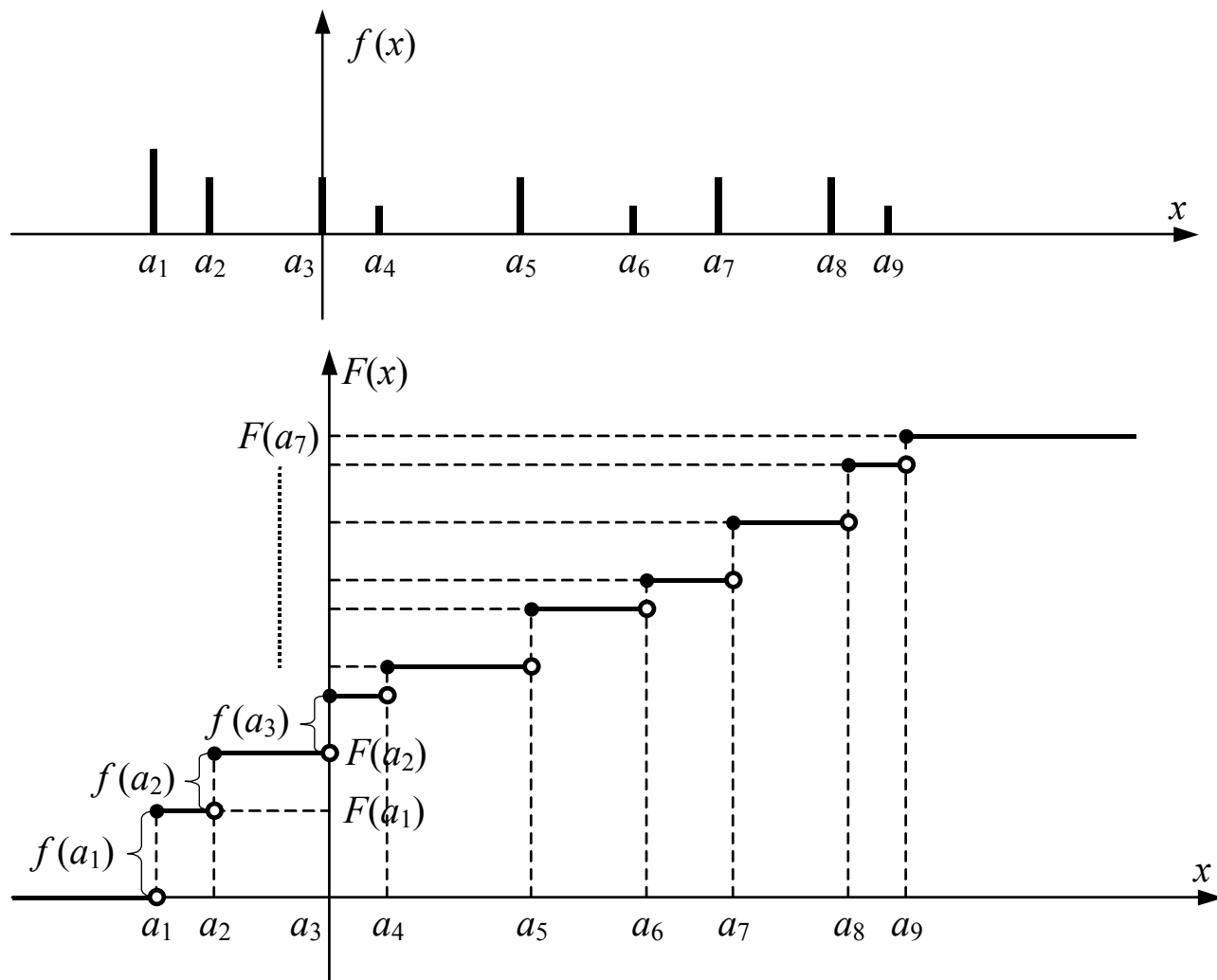
Για παράδειγμα, αν $X \in \{0,1,2,\dots\}$ τότε

$$F_X(x) = P(X \leq x) = P(X \leq [x]) = \sum_{i=0}^{[x]} P(X = i) = \sum_{i=0}^{[x]} f_X(i), \quad x \in R.$$

$$f_X(i) = P(X = i) = F_X(i) - F_X(i-1), \quad i = 2,3,\dots, \quad f_X(0) = F_X(0)$$

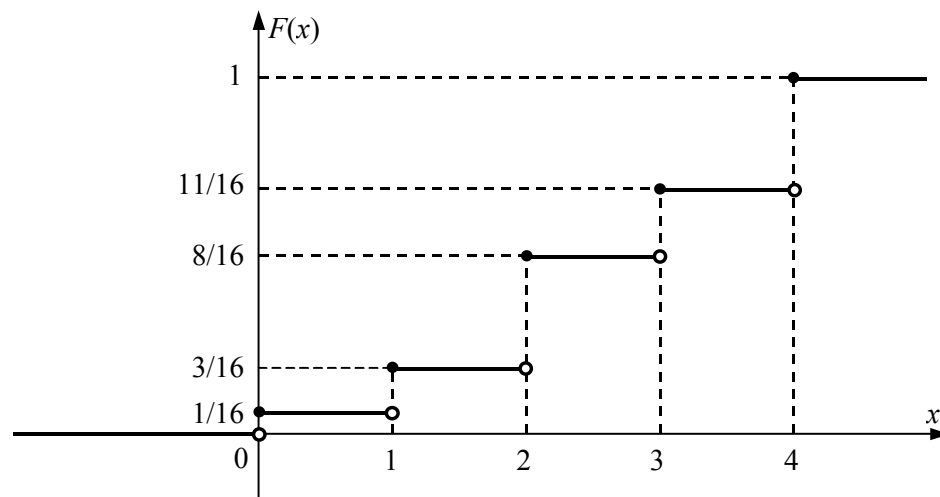
- Στις διακριτές κατανομές, η σ.κ. αρκεί να ορίζεται μόνο στα σημεία όπου αλλάζει τιμές (παρουσιάζει «άλματα»).

Γενικά, το γράφημα της συνάρτησης κατανομής $F=F_X$ και της αντίστοιχης συνάρτησης πιθανότητας $f=f_X$ μιας διακριτής τ.μ. X θα είναι της μορφής:



Παράδειγμα. Έστω ότι η τ.μ. X έχει σ.κ.

$$F(x) = \begin{cases} 0, & -\infty < x < 0 \\ 1/16, & 0 \leq x < 1 \\ 3/16, & 1 \leq x < 2 \\ 1/2, & 2 \leq x < 3 \\ 11/16, & 3 \leq x < 4 \\ 1, & 4 \leq x < \infty. \end{cases}$$



Να υπολογισθούν οι $P(1 < X \leq 3)$, $P(X > 2)$, $P(1 \leq X < 4)$ και $f(3)$

Λύση. Θα είναι,

$$P(1 < X \leq 3) = F(3) - F(1) = \frac{11}{16} - \frac{3}{16} = \frac{1}{2}, \quad P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = \frac{1}{2},$$

και

$$P(1 \leq X < 4) = P(0 < X \leq 3) = F(3) - F(0) = \frac{10}{16}, \quad f_X(3) = P(X = 3) = F(3) - F(2) = \frac{3}{16}$$

Άσκηση Να βρεθεί η τιμή της σταθεράς c , έτσι ο παρακάτω τύπος να ορίζει μια συνάρτηση πιθανότητας

$$f(x) = c3^{x-1} / 4^x, R_X = \{1,2,3,\dots\}$$

Να βρεθεί η αντίστοιχη συνάρτηση κατανομής.

Λύση. Θα πρέπει $f \geq 0$ και οι τιμές της να έχουν άθροισμα 1, δηλαδή¹

$$\sum_{x=1}^{\infty} f(x) = 1 \Leftrightarrow \sum_{x=1}^{\infty} c \frac{3^{x-1}}{4^x} = \frac{c}{4} \sum_{x=1}^{\infty} \left(\frac{3}{4}\right)^{x-1} = \frac{c}{4} \sum_{x=0}^{\infty} \left(\frac{3}{4}\right)^x = \frac{c}{4} \frac{1}{1 - \frac{3}{4}} = c \Rightarrow c = 1$$

Η αντίστοιχη συνάρτηση κατανομής θα είναι

$$F(y) = \sum_{x=1}^y f(x) = \sum_{x=1}^y \frac{3^{x-1}}{4^x} = \frac{1}{4} \sum_{x=0}^{y-1} \left(\frac{3}{4}\right)^x = \frac{1}{4} \frac{1 - \left(\frac{3}{4}\right)^y}{1 - \frac{3}{4}} = 1 - \left(\frac{3}{4}\right)^y, \quad y = 1, 2, \dots$$

και $F(y) = 0$ για $y < 1$.

¹ Υπενθυμίζονται ορισμένες βασικές σχέσεις

$$\sum_{x=1}^n x = \frac{n(n+1)}{2}, \quad \sum_{x=1}^n x^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{x=0}^n a^x = \frac{1-a^{n+1}}{1-a}, a \neq 1, \quad \sum_{x=0}^{\infty} a^x = \lim_{n \rightarrow \infty} \sum_{x=1}^n a^x = \frac{1}{1-a}, |a| < 1$$

Συνεχείς τυχαίες μεταβλητές.

Παράδειγμα Έστω X ο χρόνος ζωής ενός λαμπτήρα. Διαισθητικά, αναμένουμε ότι,

$$P(x - h < X \leq x) \rightarrow 0 \quad \text{όταν} \quad h \rightarrow 0 ,$$

(θα πρέπει η $P(X = x)$ να γίνεται ακριβώς μηδέν διότι το ενδεχόμενο να χαλάσει ο λαμπτήρας τη δεδομένη **στιγμή** x φαίνεται απίθανο διαισθητικά)

- Μπορούμε λοιπόν να υποθέσουμε ότι υπάρχει μία θετική συνάρτηση $f(x)$ για την οποία θα ισχύει ότι

$$P(x - h < X \leq x) \approx f(x)h \quad \text{για } h \text{ πολύ μικρό,}$$

ή πιο αυστηρά,

$$\frac{F(x) - F(x - h)}{h} = \frac{P(x - h < X \leq x)}{h} \rightarrow f(x) \quad \text{για } h \rightarrow 0.$$

- Με άλλα λόγια, αναμένουμε ότι,

$$\frac{d}{dx} F_X(x) = f(x),$$

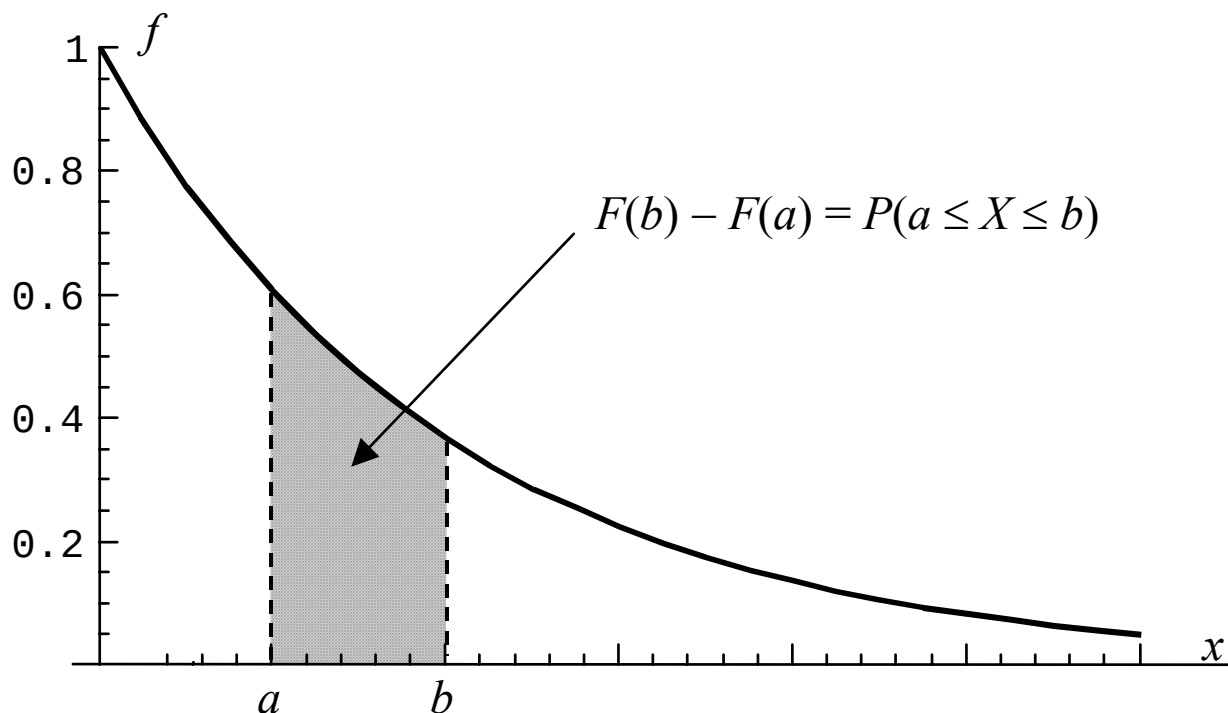
για κάποια θετική συνάρτηση $f(x)$.

- Ολοκληρώνοντας την παραπάνω από το a έως το b προκύπτει τελικά ότι

$$F_X(b) - F_X(a) = \int_a^b f(x) dx$$

για κάποια θετική συνάρτηση f .

- Δηλαδή η F_X θα πρέπει να υπολογίζεται με βάση το εμβαδόν κάτω από μία καμπύλη f_X όπως π.χ. στο παρακάτω σχήμα:



- Η κατηγορία των τ.μ. με κατανομές που περιγράφονται με αυτό τον τρόπο έχουν ιδιαίτερο ενδιαφέρον στις εφαρμογές. Συγκεκριμένα θα έχουμε τον ακόλουθο ορισμό

Μία τ.μ. X θα καλείται **συνεχής** αν υπάρχει μία μη αρνητική συνάρτηση f_X ώστε

$$F_X(b) - F_X(a) = \int_a^b f_X(x) dx, \quad a, b \in \mathbb{R}$$

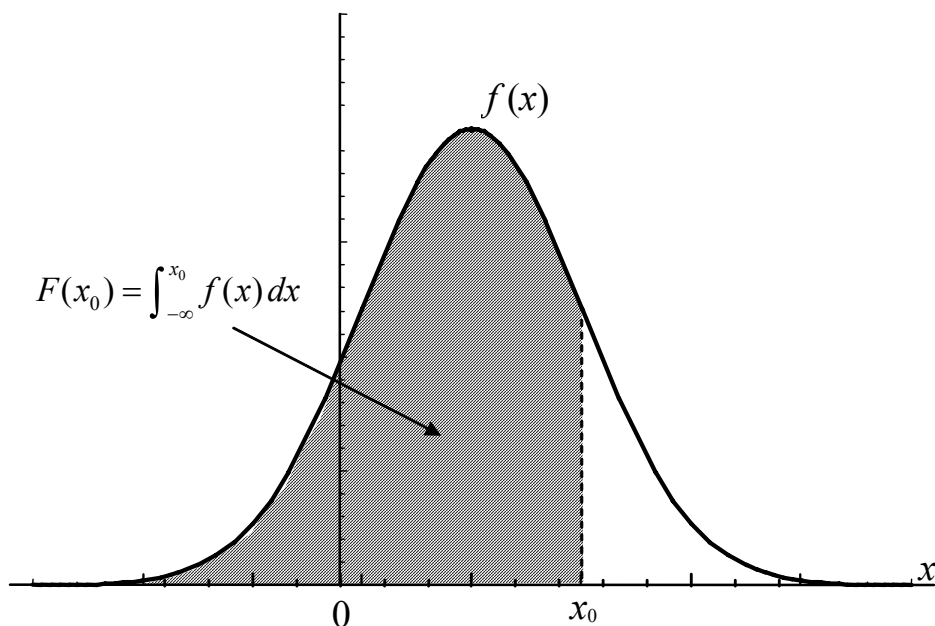
Η συνάρτηση f_X θα καλείται **συνάρτηση πυκνότητας πιθανότητας** της συνεχούς τ.μ. X .

- Εδώ, η συνολική πιθανότητα «απλώνεται» με συνεχή τρόπο πάνω σε όλα τα σημεία ενός διαστήματος του $\mathbb{R}$ (ή ακόμη και σε ολόκληρο το $\mathbb{R}$) όπως π.χ. κατανέμεται μία μάζα κατά μήκος μιας ράβδου
 - ειδικότερα, η f δείχνει την «πυκνότητα» με την οποία έχει «απλωθεί» η συνολική πιθανότητα στον $\mathbb{R}$
- Η περιγραφή της κατανομής γίνεται και πάλι μέσω της σ.κ. $F(x) = P(X \leq x)$ η οποία τώρα είναι συνεχής (και μάλιστα παραγωγίσιμη) συνάρτηση και δεν παρουσιάζει «άλλματα» όπως στην διακριτή περίπτωση.
- Τώρα δεν μπορούμε να χρησιμοποιήσουμε την $P(X=x)$ διότι η πιθανότητα αυτή είναι τώρα πάντοτε ίση με 0.

- Οι τύποι που ισχύουν για τη σ.π.π. είναι αντίστοιχοι με αυτούς που ισχύουν για τη σ.π. (στη θέση αθροισμάτων και διαφορών τώρα έχουμε ολοκληρώματα και παράγωγους)

$$F_X(x) = F_X(x) - F_X(-\infty) = \int_{-\infty}^x f_X(t) dt. \quad \frac{d}{dx} F_X(x) = f_X(x)$$

- Επίσης, θα ισχύει ότι $\int_{-\infty}^{\infty} f_X(x) dx = F_X(\infty) - F(-\infty) = 1 - 0 = 1$.
- Η $F(x)$ ισούται με το εμβαδόν κάτω από την σ.π.π. f από το $-\infty$ έως το x . Σχηματικά:



Άσκηση. Έστω συνεχής τ.μ. X η οποία παίρνει τιμές στο διάστημα $[1,3]$ με πυκνότητα

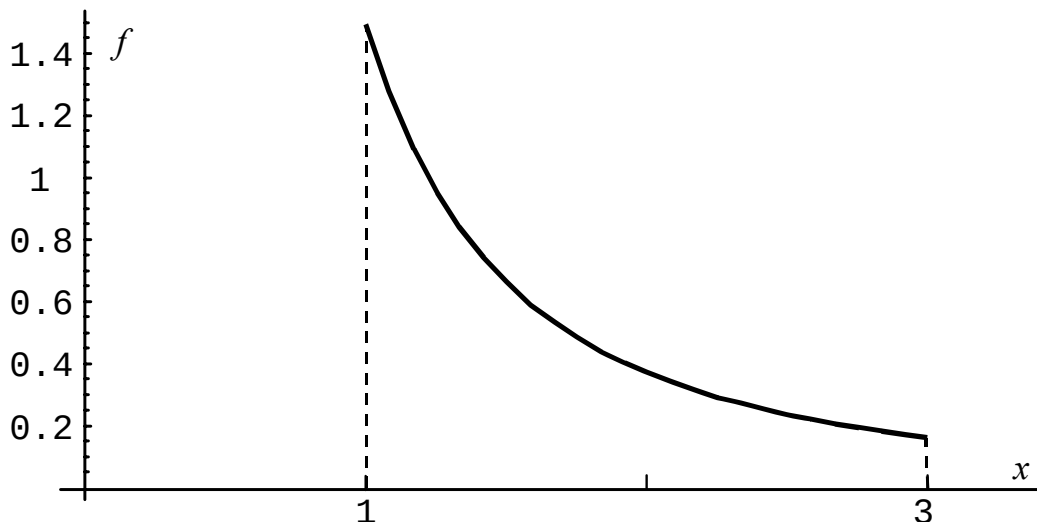
$$f(x) = \frac{a}{x^2}.$$

Να υπολογίσετε: (1) τη σταθερά a , (2) τη σ.κ. F , και (3) την πιθανότητα $P(X > 2)$.

Λύση. Η $f=0$ στο $\mathbb{R} \setminus [1,3]$ θα είναι 0. Θα πρέπει να ισχύει ότι $\int_{-\infty}^{\infty} f(x)dx = 1$. Αλλά,

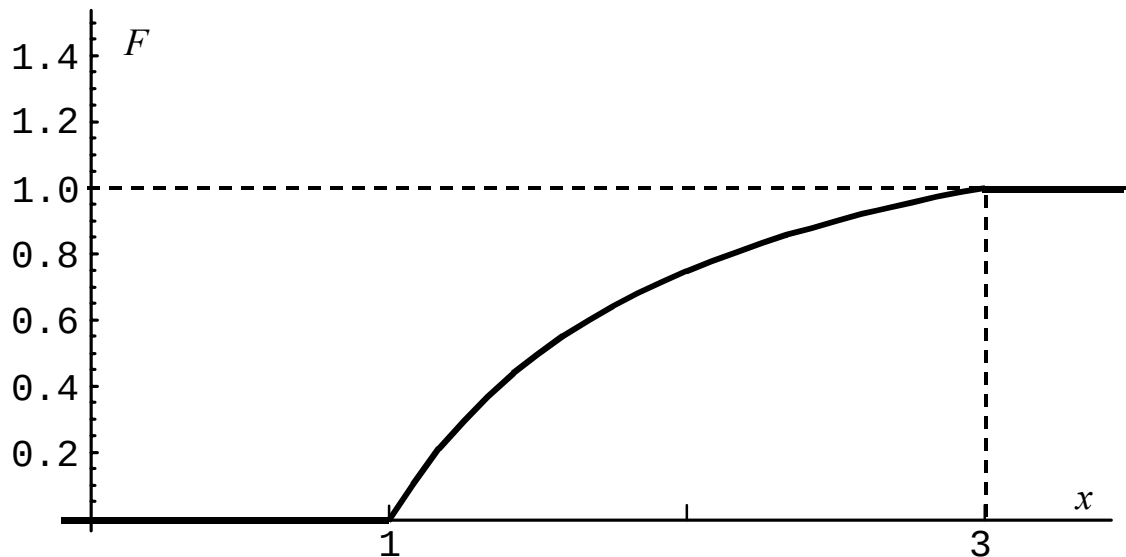
$$\int_{-\infty}^{\infty} f(x)dx = a \int_1^3 x^{-2} dx = a \int_1^3 x^{-2} dx = a \left(\frac{x^{-1}}{-1} \right)_1^3 = a \left(\frac{3^{-1}}{-1} - \frac{1^{-1}}{-1} \right) = a \left(1 - \frac{1}{3} \right) = \frac{2a}{3}$$

και επομένως, $a = 3/2$.



2) Η συνάρτηση κατανομής θα είναι ίση με

$$F(x) = \int_{-\infty}^x f(t)dt = \begin{cases} \int_{-\infty}^x 0dt = 0, & x < 1 \\ \int_{-\infty}^1 0dt + \int_1^x \frac{3}{2t^2} dt = 0 + \frac{3}{2} \int_1^x t^{-2} dt = \frac{3}{2} (-t^{-1})_1^x = \frac{3(x-1)}{2x}, & 1 \leq x \leq 3 \\ \int_{-\infty}^1 0dt + \int_1^3 \frac{3}{2t^2} dt + \int_3^x 0dt = 0 + 1 + 0 = 1, & x > 3 \end{cases}$$



3) Θα ισχύει ότι

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - \frac{3(2-1)}{2 \cdot 2} = \frac{1}{4}.$$

Συναρτήσεις μιας τυχαίας μεταβλητής.

- Σε αρκετές περιπτώσεις εμφανίζεται η ανάγκη μελέτης μιας συνάρτησης

$$Y = g(X)$$

μιας τ.μ. X . Η συνάρτηση αυτή θα είναι μία νέα τ.μ (π.χ. X^2 , e^{aX} , $\sqrt{X}$, $1/X$ κ.ο.κ.)

- Ποια θα είναι η σ.κ. F_Y της τ.μ. $Y = g(X)$ αν είναι γνωστή η σ.κ. F_X της τ.μ. X ;
- Αν η g είναι αύξουσα (και άρα αντιστρέψιμη) στο πεδίο τιμών της X , θα έχουμε ότι

$$F_Y(x) = P(Y \leq x) = P(g(X) \leq x) = P(X \leq g^{-1}(x)) = F_X(g^{-1}(x)),$$

(αν η g είναι φθίνουσα αλλάζει η $\leq$) και αν π.χ. η X είναι συνεχής τ.μ., τότε

$$f_Y(x) = \frac{d}{dx} F_Y(x) = \frac{d}{dx} F_X(g^{-1}(x)) = f_X(g^{-1}(x)) \frac{dg^{-1}(x)}{dx}.$$

• **Παράδειγμα.** Έστω X συνεχής τ.μ. με τιμές στο $[0, \infty)$ και

$$Y = \alpha X^2 + b, \quad \alpha > 0, b \in \mathbb{R}$$

Τότε

$$F_Y(x) = P(Y \leq x) = P(\alpha X^2 + b \leq x) = P(X \leq \sqrt{\frac{x-b}{\alpha}}) = F_X(\sqrt{\frac{x-b}{\alpha}}), \quad x \geq b,$$

και $F_Y(x) = 0, x < b$ ενώ

$$f_Y(x) = \frac{d}{dx} F_Y(x) = f_X(g^{-1}(x)) \frac{dg^{-1}(x)}{dx} = f_X\left(\sqrt{\frac{x-b}{\alpha}}\right) \frac{1}{2\alpha\sqrt{\frac{x-b}{\alpha}}} \text{ για } x \geq b.$$

- Αν π.χ. $f_X(x) = 1, x \in [0, 1]$ και $f_X(x) = 0, x \notin [0, 1]$, τότε η νέα τ.μ. $Y = X^2 \in [0, 1]$ έχει σ.π.π. την

$$f_Y(x) = f_X(\sqrt{x}) \frac{1}{2\sqrt{x}} = \frac{1}{2\sqrt{x}}, \quad x \in [0, 1].$$