ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ ΤΗΣ ΜΟΡΦΗΣ $\mathbf{Y = g(X_1, X_2, \ldots, X_n)}$

Έστω n τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$, οι οποίες έχουν κάποια από κοινού κατανομή. Ας υποθέσουμε ότι επιθυμούμε να προσδιορίσουμε την κατανομή της τυχαίας μεταβλητής $Y = g(X_1, X_2, \ldots, X_n)$. Η θεωρία των ένα-προς-ένα μετασχηματισμών δεν ισχύει εδώ, γιατί ενδέχεται δύο διαφορετικές τιμές του διανύσματος $(x_1, x_2, \ldots, x_n)$ να οδηγούν στην ίδια τιμή $y = g(x_1, x_2, \ldots, x_n)$ της μεταβλητής Y. Η μέθοδος, που συχνά χρησιμοποιείται στην περίπτωση αυτή για τον προσδιορισμό της κατανομής της τυχαίας μεταβλητής Y, στηρίζεται στην θεώρηση n-1 επιπλέον μετασχηματισμών

$$y_i = g_i(x_1, x_2, \ldots, x_n), \quad i=1,2,\ldots,n-1,$$

οι οποίοι μαζί με τον μετασχηματισμό $y_n = g(x_1, x_2, \ldots, x_n)$ οδηγούν σε έναν ένα-προς-ένα μετασχηματισμό από το διάνυσμα $(x_1, x_2, \ldots, x_n)$ στο διάνυσμα $(y_1, y_2, \ldots, y_n)$. Τότε, με βάση τα όσα αναπτύχθηκαν στα προηγούμενα και ανάλογα με τον εάν οι τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ είναι διακριτές ή συνεχείς, μπορούμε να προσδιορίσουμε την από κοινού συνάρτηση πιθανότητας

$$P_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n)$$

ή την από κοινού συνάρτηση πυκνότητας πιθανότητας

$$f_{Y_1, Y_2, \ldots, Y_n}(y_1, y_2, \ldots, y_n)$$

του τυχαίου διανύσματος $Y_1, Y_2, \ldots, Y_n$. Από την προκύπτουσα από κοινού κατανομή, μπορούμε τότε να προσδιορίσουμε την περιθώρια συνάρτηση πιθανότητας $P_{Y_n}(y_n) = p_Y(y)$ ή την περιθώρια συνάρτηση πυκνότητας πιθανότητας $f_{Y_n}(y_n) = f_Y(y)$ της τυχαίας μεταβλητής Y.

Παράδειγμα: Σε προηγούμενο παράδειγμα, είχαμε θεωρήσει τις συνεχείς τυχαίες μεταβλητές X_1 και X_2, που αναφέρονταν στις ανεξάρτητες μετρήσεις των χρόνων που απαιτούσε η ολοκλήρωση δύο παρόμοιων εργασιών. Είχαμε υποθέσει εκεί ότι η από κοινού συνάρτηση κατανομής των μεταβλητών αυτών ήταν

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

$$= \begin{cases} \dfrac{\theta^{2r}\, x_1^{r-1}\, x_2^{r-1}\, e^{-\theta x_1} e^{-\theta x_2}}{[\Gamma(r)]^2}, & \text{αν } 0 < x_1 < \infty, \\ & \quad 0 < x_2 < \infty, \\ 0, & \text{διαφορετικά.} \end{cases}$$

Ας υποθέσουμε ότι ενδιαφερόμαστε για την κατανομή του λόγου $Y = X_1 / X_2$. Για να προσδιορίσουμε την κατανομή αυτή, ας θεωρήσουμε τον επιπλέον μετασχηματισμό $Y_1 = X_1 + X_2$ και ας θυμηθούμε ότι, στο αρχικό παράδειγμα, ο διμεταβλητός μετασχηματισμός

$$Y_1 = X_1 + X_2 \ , \ \ Y_2 = \frac{X_1}{X_2},$$

ήταν ένας ένα-προς-ένα μετασχηματισμός, ο οποίος οδήγησε στην από κοινού συνάρτηση πυκνότητας πιθανότητας

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \dfrac{\theta^{2r} y_1^{2r-1} e^{-\theta y_1}}{\Gamma(2r)} \dfrac{y_2^{r-1}}{B(r,r)\,(1+y_2)^{2r}}, & \text{αν } 0 < y_1 < \infty, \\ & \quad 0 < y_2 < \infty, \\ 0, & \text{διαφορετικά.} \end{cases}$$

Η περιθώρια συνάρτηση πυκνότητας πιθανότητας $f_{Y_2}(y_2) = f_Y(y)$ της τυχαίας μεταβλητής Y μπορεί, τότε, να προσδιορισθεί ως εξής:

$$f_Y(y) = \int_0^{+\infty} f_{Y_1,Y_2}(y_1,y)\,dy_1 = \begin{cases} \dfrac{y^{r-1}}{B(r,r)(1+y)^{2r}} \displaystyle\int_0^{+\infty} \frac{\theta^{2r} y_1^{2r-1} e^{-\theta y_1}}{\Gamma(2r)}\,dy_1 = \frac{y^{r-1}}{B(r,r)(1+y)^{2r}}, & 0 < y < \infty \\ \displaystyle\int_0^{+\infty} dy_1 = 0, & y \le 0. \end{cases}$$

Στο προηγούμενο παράδειγμα, θα μπορούσαμε να ενδιαφερόμαστε να προσδιορίσουμε την κατανομή του συνολικού χρόνου που απαιτείται για την ολοκλήρωση των δύο εργασιών. Στην περίπτωση αυτή, θα μπορούσαμε να

προσδιορίσουμε την κατανομή της τυχαίας μεταβλητής $Y_1 = X_1 + X_2$ βρίσκοντας την περιθώρια συνάρτηση πυκνότητας πιθανότητάς της από την σχέση

$$f_{Y_1}(y_1) = \int_0^{+\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2$$

$$= \begin{cases} \dfrac{\theta^{2r} y_1^{2r-1} e^{-\theta y_1}}{\Gamma(2r)}, & \text{αν } 0 < y_1 < \infty \\ 0, & \text{διαφορετικά.} \end{cases}$$

Παρατηρούμε ότι η τυχαία μεταβλητή $Y_1 = X_1 + X_2$ έχει την κατανομή Γάμμα με παραμέτρους θ και 2r.

Παρατήρηση: Θα πρέπει να σημειωθεί εδώ ότι δεν είναι πάντα εύκολος ο καθορισμός των επιπλέον μετασχηματισμών $Y_i = g_i(X_1, X_2, \ldots, X_n)$, i=1, 2,..., n-1, οι οποίοι απαιτούνται ώστε, μαζί με τον μετασχηματισμό $Y = g(X_1, X_2, \ldots, X_n)$, να επιτυγχάνεται ένας ένα-προς-ένα μετασχηματισμός. Αν στο προηγούμενο παράδειγμα δεν είχαμε ήδη προσδιορίσει την από κοινού κατανομή των μεταβλητών $Y_1 = X_1 + X_2$ και $Y_2 = X_1 / X_2$, ενδέχεται να είχαμε κάποια δυσκολία στο να βρούμε έναν μετασχηματισμό $Y_1 = g_1(X_1, X_2)$, ο οποίος, μαζί με τον $Y_2 = X_1 / X_2$, θα οδηγούσε σε έναν κατάλληλο ένα-προς-ένα μετασχηματισμό για τον προσδιορισμό της περιθώριας κατανομής της μεταβλητής Y_2.

Ένας τρόπος, για να παρακαμφθεί αυτή η έμμεση αντιμετώπιση του προβλήματος, είναι η χρησιμοποίηση της αθροιστικής συνάρτησης κατανομής.

Εάν $Y = g(X_1, X_2, \ldots, X_n)$, τότε

$$F_Y(y) = P(Y \le y)$$

$$= P\{g(X_1, X_2, \ldots, X_n) \le y\}$$

$$= \int\cdots\int_{g(x_1,x_2,\ldots,x_n)\le y} f_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Ας υποθέσουμε ότι υπάρχει μία συνάρτηση $h(y; x_2,\ldots,x_n)$ τέτοια ώστε $g(x_1, x_2,\ldots, x_n) \le y$ τότε και μόνο τότε εάν $x_1 \le h(y; x_2,\ldots,x_n)$. Τότε,

$$\begin{aligned} F_Y(y) &= \int\cdots\int_{g(x_1,x_2,\ldots,x_n)\le y} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\,dx_1\cdots dx_n \\ &= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{h(y;x_2,\ldots,x_n)} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)dx_1\right]dx_2\cdots dx_n \\ &= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} q(y;x_2,\ldots,x_n)dx_2\cdots dx_n, \end{aligned}$$

όπου

$$q(y;x_2,\ldots,x_n) = \int_{-\infty}^{h(y;x_2,\ldots,x_n)} f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)dx_1.$$

Προφανώς

$$\frac{\partial q(y; x_2,\ldots, x_n)}{\partial y} = \frac{\partial h(y; x_2,\ldots, x_n)}{\partial y}\left[f_{X_1,X_2,\ldots,X_n}(h(y; x_2,\ldots, x_n), x_2,\ldots, x_n)\right]$$

Οι παραπάνω σχέσεις οδηγούν στην συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y:

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y}F_Y(y) \\ &= \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial\, h(y; x_2,\ldots, x_n)}{\partial y}\left[f_{X_1,X_2,\ldots,X_n}(h(y; x_2,\ldots, x_n), x_2,\ldots, x_n)\right]dx_2\cdots dx_n, \end{aligned}$$

με την προϋπόθεση ότι ισχύει

$$\frac{\partial}{\partial\, y}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} q(y; x_2,\ldots, x_n)\,dx_2\cdots dx_n = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial}{\partial\, y}q(y; x_2,\ldots, x_n)\,dx_2\cdots dx_n.$$

Βέβαια, η σχέση αυτή δεν ισχύει πάντα και απαραίτητη προϋπόθεση είναι οι συναρτήσεις $q(y;x_2,\ldots,x_n)$ και $\frac{\partial\, q(y; x_2,\ldots, x_n)}{\partial\, y}$ να είναι καλώς ορισμένες συνεχείς συναρτήσεις των $y, x_2,\ldots,x_n$ για $-\infty < y, x_2, x_3,\ldots, x_n < \infty$.

Παράδειγμα *(Η κατανομή t)*: Ένα πρόβλημα μεγάλης σημασίας στην θεωρία της στατιστικής συμπερασματολογίας είναι το εξής.

Ας υποθέσουμε ότι οι τυχαίες μεταβλητές X_1 και X_2 είναι ανεξάρτητες και ότι η X_1 ακολουθεί την τυποποιημένη κανονική κατανομή, ενώ η X_2 ακολουθεί την

κατανομή X^2 με n βαθμούς ελευθερίας. Ενδιαφερόμαστε να προσδιορίσουμε την κατανομή της τυχαίας μεταβλητής $Y = \dfrac{X_1}{\sqrt{X_2 / n}}$.

Λύση: Χρησιμοποιώντας τον συμβολισμό και τα βήματα της μεθόδου που εκθέσαμε παραπάνω, έχουμε ότι $g(x_1, x_2) = \dfrac{x_1}{\sqrt{x_2/n}}$, και ότι $\dfrac{x_1}{\sqrt{x_2 / n}} \le y$ τότε και μόνο τότε αν $x_1 \le y\sqrt{x_2 / n}$,έτσι ώστε $h(y; x_2) = y\sqrt{x_2 / n}$. Από τον ορισμό των τυχαίων μεταβλητών X_1 και X_2 έχουμε ότι

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

$$= \begin{cases} \left(\dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \right) \dfrac{x_2^{n/2-1} e^{-\frac{1}{2}x_2}}{2^{n/2}\Gamma(n/2)}, & -\infty < x_1 < \infty \\ & -\infty < x_2 < \infty \\ 0, & \text{διαφορετικά.} \end{cases}$$

Επομένως,

$$q(y; x_2) = \int_0^{y\sqrt{x_2/n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \frac{x_2^{n/2-1} e^{-\frac{1}{2}x_2}}{2^{n/2}\Gamma(n/2)} dx_1,$$

και

$$\frac{\partial}{\partial y} q(y; x_2) = \left[\sqrt{\frac{x_2}{n}} \right] \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y\sqrt{x_2/n})^2} \frac{x_2^{n/2-1} e^{-\frac{1}{2}x_2}}{2^{n/2}\Gamma(n/2)}.$$

Επειδή και οι δύο αυτές συναρτήσεις είναι καλώς ορισμένες και συνεχείς συναρτήσεις των y και x_2 για όλα τα $-\infty < y < \infty,\ 0 \le x_2 < \infty$, έπεται ότι

$$f_Y(y) = \frac{1}{\sqrt{2\pi n}\,\Gamma(n/2)\,2^{n/2}} \int_0^{+\infty} x_2^{\frac{1}{2}(n-1)} e^{-\frac{1}{2}x_2(1+y^2/n)} dx_2$$

$$= \frac{\Gamma\left[\dfrac{n+1}{2}\right]}{\Gamma(n/2)\sqrt{\pi n}\,(1 + y^2/n)^{(n+1)/2}}, \qquad -\infty < y < +\infty.$$

Η κατανομή που ορίζεται από την παραπάνω συνάρτηση πυκνότητας πιθανότητας είναι η γνωστή *κατανομή t του Student με n βαθμούς ελευθερίας*. Όπως, με άλλη

μέθοδο, μπορεί ν' αποδειχθεί (βλέπε π.χ. *Ε. Ξεκαλάκη και Ι. Πανάρετο, 1993)*, η κατανομή t μπορεί να προκύψει ως κατανομή του τυποποιημένου μέσου μίας κανονικής κατανομής στο πλαίσιο προβλημάτων στατιστικής συμπερασματολογίας. Η κατανομή αυτή είχε προταθεί από τον Αγγλο στατιστικό William Sealy Gosset (1876-1937), ο οποίος είχε χρησιμοποιήσει το ψευδώνυμο Student.

Παράδειγμα (*Η κατανομή* X^2): Έστω ότι $X_1, X_2, \ldots, X_n$ είναι n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την τυποποιημένη κανονική κατανομή, δηλαδή, $X_i \sim N(0,1)$, i=1,2,...,n. Ενδιαφερόμαστε να προσδιορίσουμε την κατανομή της τυχαίας μεταβλητής $Y = X_1^2 + X_2^2 + \ldots + X_n^2$.

Λύση: Έστω $Y_i = X_i^2$, i=1,2,...,n. Τότε, η τυχαία μεταβλητή Y μπορεί εναλλακτικά να γραφεί με τη μορφή $Y_1 + Y_2 + \ldots + Y_n$. Επομένως, η συνάρτηση κατανομής της τυχαίας μεταβλητής Y μπορεί να προκύψει από την μορφή της συνάρτησης κατανομής της τυχαίας μεταβλητής Y .

Για την τελευταία έχουμε ότι

$$F_{Y_i}(y) = P(X_i^2 \le y)$$

$$= P(|X_i| \le \sqrt{y})$$

$$= P(-\sqrt{y} \le X_i \le \sqrt{y})$$

$$= \Phi_{X_i}(\sqrt{y}) - \Phi_{X_i}(-\sqrt{y})$$

$$= 2\Phi_{X_i}(\sqrt{y}) - 1.$$

Η τελευταία ισότητα είναι αποτέλεσμα του γεγονότος ότι η τυποποιημένη κανονική κατανομή των μεταβλητών $X_1, X_2, \ldots, X_n$ είναι συμμετρική. Επομένως, η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y_i είναι η

$$f_{Y_i}(y) = f_{X_i}(\sqrt{y})\frac{1}{\sqrt{y}}$$

$$= \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}, \quad y > 0 .$$

Λαμβάνοντας υπόψη το γεγονός ότι Γ(1/2) = π, η παραπάνω συνάρτηση

πυκνότητας γράφεται με την εξής μορφή:

$$f_{Y_i}(y) = \frac{1}{2^{1/2}\,\Gamma(1/2)}\; y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}.$$

Αυτή, όμως, είναι η συνάρτηση πυκνότητας πιθανότητας της κατανομής Γάμμα με παραμέτρους $\alpha = \frac{1}{2}$ και θ = 2. Κατά συνέπεια, η τυχαία μεταβλητή Y_i είναι άθροισμα n ανεξαρτήτων Γάμμα τυχαίων μεταβλητών, κάθε μία με παραμέτρους $\alpha = \frac{1}{2}$ και θ = 2. Όπως είναι γνωστό (βλέπε π.χ. *Ε. Ξεκαλάκη και Ι. Πανάρετο, 1993),* το άθροισμα n ανεξαρτήτων και ισονόμων Γάμμα μεταβλητών με παραμέτρους α και θ αποτελεί τυχαία μεταβλητή η οποία ακολουθεί την κατανομή Γάμμα με παραμέτρους nα και θ. Επομένως, η τυχαία μεταβλητή

$$\begin{aligned} Y &= Y_1 + Y_2 + \ldots + Y_n \\ &= X_1^2 + X_2^2 + \ldots + X_n^2 \end{aligned}$$

ακολουθεί την κατανομή Γάμμα με παραμέτρους α = n/2 και θ = 2, δηλαδή την κατανομή X^2 με n βαθμούς ελευθερίας.

Θα πρέπει να σημειωθεί εδώ ότι η κατανομή X^2 ανακαλύφθηκε για πρώτη φορά από τον Γερμανό φυσικό F. A. Helmert (1843-1917). Αργότερα, ο Karl Pearson (1857-1936), ένας Αγγλος στατιστικός, κατέληξε στην κατανομή αυτή στο πλαίσιο της ερευνητικής του εργασίας σχετικής με την στατιστική ανάλυση πινάκων συναφείας. Η κατανομή αυτή επίσης χρησιμοποιείται σε σχέση με προβλήματα στατιστικής συμπερασματολογίας για την διασπορά ενός κανονικού πληθυσμού.

Παράδειγμα (*Η κατανομή F*): Έστω ότι οι τυχαίες μεταβλητές X_1 και X_2 είναι ανεξάρτητες X^2 μεταβλητές. Πιο συγκεκριμένα, υποθέτουμε ότι $X_1 \sim X^2_{n_1}$ και $X_2 \sim X^2_{n_2}$. Ενδιαφερόμαστε να προσδιορίσουμε την κατανομή της τυχαίας μεταβλητής $Y = \frac{X_1/n_1}{X_2/n_2}$.

Λύση: Η κατανομή της τυχαίας μεταβλητής Y μπορεί να προσδιορισθεί έμμεσα, χρησιμοποιώντας έναν ένα-προς-ένα μετασχηματισμό, όπως, για παράδειγμα, τον μετασχηματισμό $Y_1 = Y = (X_1/n_1)/(X_2/n_2)$, $Y_2 = X_1 + X_2$. Αφού προσδιορίσουμε την από κοινού συνάρτηση πυκνότητας πιθανότητας $f_{Y_1,Y_2}(y_1, y_2)$ των τυχαίων

μεταβλητών Y_1, Y_2, η κατανομή της τυχαίας μεταβλητής $Y=Y_1$ προκύπτει αμέσως από την σχέση $f_Y(y) = f_{Y_1}(y) = \int_0^{+\infty} f_{Y_1,Y_2}(y, y_2)\, dy_2$.

Εναλλακτικά, θα μπορούσαμε να χρησιμοποιήσουμε την έμμεση μέθοδο που στηρίζεται στην συνάρτηση κατανομής της τυχαίας μεταβλητής Y.

Χρησιμοποιώντας το συμβολισμό που εισήχθη στο πλαίσιο της μεθόδου αυτής, έχουμε ότι $g(x_1, x_2) = \frac{x_1/n_1}{x_2/n_2} \le y$ τότε και μόνο τότε αν $x_1 \le y\left[\frac{n_1}{n_2}\right] x_2$, έτσι ώστε $h(y; x_2) = y(n_1/n_2)x_2$. Από τον ορισμό των τυχαίων μεταβλητών X_1 και X_2 προκύπτει ότι

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2)$$

$$= \begin{cases} \dfrac{x_1^{\frac{1}{2}n_1 - 1} e^{-\frac{1}{2}x_1}}{2^{\frac{1}{2}n_1}\,\Gamma(n_1/2)} \dfrac{x_2^{\frac{1}{2}n_2 - 1} e^{-\frac{1}{2}x_2}}{2^{\frac{1}{2}n_2}\,\Gamma(n_2/2)}, & -\infty \le x_1 x_2 \le \infty \\ 0, & \text{διαφορετικά} \end{cases}$$

Επομένως, οι συναρτήσεις

$$q(y; x_2) = \int_0^{y(n_1/n_2)x_2} \frac{x_1^{\frac{1}{2}n_1 - 1} x_2^{\frac{1}{2}n_2 - 1} e^{-\frac{1}{2}(x_1 + x_2)}}{2^{\frac{1}{2}(n_1+n_2)}\,\Gamma(n_1/2)\,\Gamma(n_2/2)}\, dx_1$$

και

$$\frac{\partial}{\partial y} q(y; x_2) = \frac{y^{\frac{1}{2}n_1 - 1} (n_1/n_2)^{\frac{1}{2}n_1} x_2^{\frac{1}{2}(n_1+n_2)-1} e^{-\frac{1}{2}x_2(1 + n_1 y/n_2)}}{2^{\frac{1}{2}(n_1+n_2)}\,\Gamma(n_1/2)\,\Gamma(n_2/2)}$$

είναι καλώς ορισμένες και συνεχείς συναρτήσεις των y και x_2, για όλα τα $0 \le y,\ x_2 < \infty$. Τότε, για $y \ge 0$,

$$f_Y(y)=\frac{(n_1/n_2)^{\frac{1}{2}n_1}\, y^{\frac{1}{2}n_1-1}}{2^{\frac{1}{2}(n_1+n_2)}\,\Gamma(n_1/2)\,\Gamma(n_2/2)}\int_0^{+\infty} x_2^{\frac{1}{2}(n_1+n_2)-1}\, e^{-\frac{1}{2}x_2(1+n_1y/n_2)}\, dx_2$$

$$=\frac{(n_1/n_2)^{\frac{1}{2}n_1}\,\Gamma((n_1+n_2)/2)\,y^{\frac{1}{2}n_1-1}}{\Gamma(n_1/2)\,\Gamma(n_2/2)\,(1+n_1y/n_2)^{\frac{1}{2}(n_1+n_2)}},$$

ενώ $f_Y(y)=0$, για $y<0$.

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y ορίζει την κατανομή που είναι γνωστή στην βιβλιογραφία ως κατανομή F με n_1 και n_2 βαθμούς ελευθερίας. Η κατανομή αυτή χρησιμοποιήθηκε από τον Snedecor στα πλαίσια της στατιστικής μεθοδολογίας η οποία είναι γνωστή ως ανάλυση διασποράς. Ο Snedecor ονόμασε την κατανομή αυτή F προς τιμή του R. A. Fisher.

ΑΘΡΟΙΣΜΑΤΑ ΚΑΙ ΣΤΑΘΜΙΣΜΕΝΑ ΑΘΡΟΙΣΜΑΤΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ

Ίσως οι πιο συχνά χρησιμοποιούμενοι μετασχηματισμοί n τυχαίων μεταβλητών X_1, $X_2,...,X_n$ της μορφής $Y = g(X_1, X_2,...,X_n)$ είναι τα σταθμισμένα αθροίσματα $Y = \sum_{i=1}^{n} \alpha_i X_i$ των μεταβλητών αυτών. Οι εφαρμογές των αθροισμάτων αυτών είναι πάρα πολλές. Ενδεικτικά αναφέρουμε τις εξής περιπτώσεις:

(α) Έστω ότι $X_1, X_2,...,X_n$ είναι n ανεξάρτητες τυχαίες μεταβλητές. Τότε, σε σχέση με πολλά πρακτικά προβλήματα, ενδιαφερόμαστε για τον μέσο $\overline{Y} = \frac{X_1 + X_2 + ... + X_n}{n}$ των μεταβλητών αυτών.

(β) Έστω ότι X_1, X_2 είναι τυχαίες μεταβλητές που παριστάνουν μετρήσεις πάνω σε κάποια χαρακτηριστικά. Π.χ. X_1 μπορεί να παριστάνει το επίπεδο λιθίου στο αίμα κάποιου ασθενούς, ενώ X_2 παριστάνει το επίπεδο λιθίου στο αίμα του ιδίου ασθενούς μετά από κάποια θεραπευτική αγωγή. Τότε, η διαφορά $Y = X_2 - X_1$ είναι κεντρικής σημασίας στο πλαίσιο του προβλήματος αυτού, αφού παρέχει ένα μέτρο της μεταβολής που προκάλεσε η θεραπευτική αγωγή.

(γ) Έστω ότι X_1, X_2, X_3, X_4 είναι τυχαίες μεταβλητές που παριστάνουν τους βαθμούς που ένας φοιτητής πέτυχε όσο αφορά κάποιο μάθημα σε δύο προόδους (X_1, X_2), μία εργασία (X_3) και στην τελική εξέταση (X_4). Ο τελικός βαθμός στο μάθημα αυτό θα μπορούσε να προσδιορίζεται από ένα άθροισμα της μορφής $Y = \sum_{i=1}^{4} X_i$ ή από ένα σταθμισμένο άθροισμα της μορφής $Y = 0.2X_1 + 0.2X_2 + 0.3X_3 + 0.3X_4$.

Όπως είναι γνωστό (βλέπε π.χ. *Ε. Ξεκαλάκη και Ι. Πανάρετο, 1993*), η κατανομή ενός σταθμισμένου αθροίσματος $Y = \sum_{i=1}^{n} \alpha_i X_i$ n τυχαίων μεταβλητών X_1, $X_2,...,X_n$ μπορεί να μελετηθεί με τη χρήση της έννοιας των ροπογεννητριών. Στην παρούσα ενότητα, θα εξετάσουμε το ίδιο πρόβλημα στο πλαίσιο της θεωρίας των ένα-προς-ένα μετασχηματισμών:

Έστω ότι τουλάχιστον ένας από τους συντελεστές α_i είναι διάφορος του μηδενός. Εστω, δηλαδή, ότι υπάρχει τουλάχιστον ένας δείκτης i για τον οποίο ισχύει $\alpha_i \neq 0$

και για απλότητα, ας υποθέσουμε ότι αυτό συμβαίνει για i = 1, δηλαδή ότι $\alpha_1 \neq 0$. Στην συνέχεια, θεωρούμε τον ένα-προς-ένα μετασχηματισμό

$$Y_1 = Y = \sum_{i=1}^{n} \alpha_i X_i, \quad Y_2 = X_2, ..., Y_n = X_n,$$

για τον οποίο, προφανώς, ισχύει ότι

$$x_i = h_i(y_1, y_2, ...y_n) = y_i, \quad i = 1,2,\ldots n, \text{ και}$$

$$x_1 = h_1(y_1, y_2, ...y_n) = \frac{y_1 - \sum_{i=2}^{n} \alpha_i y_i}{\alpha_1} = \frac{y - \sum_{i=2}^{n} \alpha_i y_i}{\alpha_1}$$

Μπορούμε, επομένως, να προσδιορίσουμε την από κοινού κατανομή των τυχαίων μεταβλητών $Y_1, Y_2, ..., Y_n$ και από αυτήν να προσδιορίσουμε την περιθώρια κατανομή της τυχαίας μεταβλητής $Y_1 = Y$. Αρα, όταν οι τυχαίες μεταβλητές $X_1, X_2, ..., X_n$ είναι διακριτές με από κοινού συνάρτηση πιθανότητας $P_{X_1, X_2, ..., X_n}(x_1, x_2, ..., x_n) = P(x_1, x_2, ..., x_n)$, τα παραπάνω βήματα οδηγούν στην εξής συνάρτηση κατανομής της τυχαίας μεταβλητής Y:

$$P_Y(y) = \sum_{y_2} \cdots \sum_{y_n} P\left[\frac{y - \sum_{i=2}^{n} \alpha_i y_i}{\alpha_1}, y_2, ..., y_n \right].$$

Με όμοιο τρόπο, όταν $X_1, X_2, ..., X_n$ είναι συνεχείς τυχαίες μεταβλητές με από κοινού συνάρτηση πυκνότητας πιθανότητας $f_{X_1, X_2, ..., X_n}(x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n)$, τότε η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y δίνεται από την σχέση

$$f_Y(y) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{1}{|\alpha_1|} f\left[\frac{y - \sum_{i=2}^{n} \alpha_i y_i}{\alpha_1}, y_2 ... y_n \right] dy_2 ... dy_n.$$

Σημείωση: Η υπόθεση ότι $\alpha_i \neq 0$, η οποία έγινε για απλότητα δεν βλάπτει τη γενικότητα. Πράγματι, αν υποθέσουμε ότι $\alpha_j \neq 0$, για κάποια τιμή $j \neq 1$, τότε η ίδια επιχειρηματολογία θα οδηγήσει στις εξής εναλλακτικές εκφράσεις για την συνάρτηση πιθανότητας και την συνάρτηση πυκνότητας πιθανότητας αντίστοιχα της μεταβλητής Y:

$$P_Y(y) = \sum_{y_1} \cdots \sum_{y_{j-1}} \sum_{y_{j+1}} \cdots \sum_{y_n} P\left[y_1, \ldots, y_{j-1}, \frac{y - \sum_{i \neq j} \alpha_i y_i}{\alpha_j}, y_{j+1}, \ldots, y_n \right]$$

$$f_Y(y) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{1}{|\alpha_j|} f\left[y_1, \ldots, y_{j-1}, \frac{y - \sum_{i \neq j} \alpha_i y_i}{\alpha_j}, y_{j+1}, \ldots, y_n \right] dy_1 \ldots dy_{j-1}\, dy_{j+1} \ldots dy_n$$

Παράδειγμα: Έστω ότι το διάνυσμα (X_1, X_2) έχει την από κοινού συνάρτηση πιθανότητας

$$P_{X_1,X_2}(x_1, x_2) = \frac{n!}{x_1! x_2! (n - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - p_1 - p_2)^{n - x_1 - x_2}$$

για κάθε $x_1, x_2 = 0,1, \ldots,n$, $0 \le x_1 + x_2 \le n$ και $P_{X_1X_2}(x_1, x_2) = 0$, διαφορετικά. Τότε, σύμφωνα με τα προηγούμενα, η συνάρτηση πιθανότητας της τυχαίας μεταβλητής $Y = X_1 + X_2$ είναι

$$P_Y(y) = \sum_{y_2=0}^{n} P_{X_1,X_2}(y - y_2, y_2) = \sum_{y_2=0}^{y} P_{X_1,X_2}(y - y_2, y_2)$$

για κάθε $y = 0,1,\ldots,n$ και $P_Y(y) = 0$, διαφορετικά. Εχουμε, όμως, ότι

$$\sum_{y_2=0}^{y} P_{X_1,X_2}(y - y_2, y_2) = \frac{n!\,(1 - p_1 - p_2)^{n-y}}{(n - y!)y!} \left[\sum_{y_2=0}^{y} \frac{y!}{y_2!(y - y_2)!} p_1^{y - y_2} p_2^{y_2} \right]$$

$$= \begin{bmatrix} n \\ y \end{bmatrix} (1 - p_1 - p_2)^{n-y} (p_1 + p_2)^y .$$

Επομένως,

$$P_Y(y) = \begin{cases} \begin{bmatrix} n \\ y \end{bmatrix} (p_1 + p_2)^y (1 - p_1 - p_2)^{n-y}, & \text{αν } y = 0,1,\ldots n, \\ 0, & \text{διαφορετικά} \end{cases}$$

η οποία είναι η συνάρτηση πιθανότητας της διωνυμικής κατανομής με παραμέτρους n και $p = p_1 + p_2$.

Παράδειγμα: Έστω X_1 και X_2 ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με παράμετρο $\theta=1$ και με από κοινού συνάρτηση πυκνότητας πιθανότητας

$$f_{X_1,X_2}(x_1,x_2)=\begin{cases} e^{-(x_1+x_2)}, & \text{αν} \quad x_1>0,\ x_2>0, \\ \\ 0 & \text{διαφορετικά} \end{cases}$$

Να προσδιορισθεί η κατανομή της τυχαίας μεταβλητής $Y = 2X_1 + 3X_2$.

Λύση: Σύμφωνα με τα προηγούμενα, η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y είναι

$$f_Y(y)=\begin{cases} \int_{-\infty}^{+\infty}\left|\frac{1}{2}\right| f_{x_1,x_2}\left(\frac{y-3y_2}{2},y_2\right)\,dy_2 & \text{αν} \quad y>0 \\ \\ 0, & \text{διαφορετικά} \end{cases}$$

Δηλαδή, για $y > 0$,

$$f_Y(y)=\frac{1}{2}\int_0^{y/3}\exp\left(-\frac{y-3y_2}{2}-y_2\right)dy_2+\frac{1}{2}\int_{y/3}^{\infty}0\,dy_2$$

$$=\frac{1}{2}\exp\left(-\frac{1}{2}y\right)\int_0^{y/3}\exp\left(-\frac{1}{2}y_2\right)dy_2$$

$$=\frac{1}{2}\exp\left(-\frac{y}{2}\right)\left[2\exp\left(\frac{y}{6}\right)-2\right]$$

$$=\exp\left(-\frac{y}{3}\right)-\exp\left(-\frac{y}{2}\right)$$

Παρατήρηση : Μερικές φορές, αντί της μεθόδου που αναπτύχθηκε στην ενότητα αυτή για την εύρεση της κατανομής σταθμισμένων αθροισμάτων τυχαίων μεταβλητών, χρησιμοποιείται μία διαφορετική μέθοδος η οποία είναι περισσότερο άμεση με την έννοια ότι δεν απαιτείται ο καθορισμός επιπλέον τυχαίων μεταβλητών για την κατασκευή ενός ένα-προς-ένα μετασχηματισμού. Η μέθοδος αυτή εφαρμόζεται στο παράδειγμα που ακολουθεί.

Παράδειγμα: Ας υποθέσουμε ότι η από κοινού κατανομή των τυχαίων μεταβλητών X_1, X_2 και X_3 ορίζεται από την εξής από κοινού συνάρτηση πιθανότητας:

(x_1, x_2, x_3)	(0,0,0)	(1,0,0)	(1,1,0)	(1,0,1)	(0,1,0)	(0,1,1)	(0,0,1)	(1,1,1)

$P(x_1, x_2, x_3)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$

Για την τυχαία μεταβλητή $Y = \frac{1}{3}(X_1 + X_2 + X_3)$ έχουμε ότι

$$P(Y = 0) = P(0,0,0) = \frac{1}{8},$$

$$P(Y = 1) = P(1,1,1) = \frac{2}{8},$$

$$P(Y = \frac{1}{3}) = P(1,0,0) + P(0,1,0) + P(0,0,1) = \frac{1}{4} + \frac{1}{16} + \frac{1}{16} = \frac{3}{8},$$

$$P(Y = \frac{2}{3}) = P(1,1,0) + P(1,0,1) + P(0,1,1) = \frac{1}{16} + \frac{1}{8} + \frac{1}{16} = \frac{2}{8}.$$

Επομένως, η συνάρτηση πιθανότητας της τυχαίας μεταβλητής Y είναι

y	0	$\frac{1}{3}$	$\frac{2}{3}$	1
P(Y=y)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{2}{8}$

ΑΘΡΟΙΣΜΑΤΑ ΤΥΧΑΙΩΝ ΜΕΤΑΒΛΗΤΩΝ - ΣΥΝΕΛΙΞΗ (CONVOLUTION)

Είδαμε στα προηγούμενα ένα παράδειγμα προσδιορισμού της κατανομής του αθροίσματος $Y = \sum_{i=1}^{n} X_i$ n τυχαίων μεταβλητών στην περίπτωση που n=2. Η γενική περίπτωση (n>2) αντιμετωπίζεται όπως και η περίπτωση των σταθμισμένων αθροισμάτων όταν $\alpha_i = 1$, i=1,2,...,n. Οι αντίστοιχες εξισώσεις για τη συνάρτηση πιθανότητας (όταν η Y είναι διακριτή) και για την συνάρτηση πυκνότητας πιθανότητας (όταν η Y είναι συνεχής) παίρνουν τη μορφή

$$P_Y(y) = \sum_{y_2} \cdots \sum_{y_n} P\left(y - \sum_{i=2}^{n} y_i, y_2, \ldots, y_n \right)$$

και

$$f_Y(y) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f\left(y - \sum_{i=2}^{n} y_i, y_2, ..., y_n \right) dy_2 ... dy_n .$$

Θα έπρεπε ίσως να σημειωθεί εδώ ότι, λόγω συμμετρίας, ισχύουν ανάλογοι τύποι στους οποίους το άθροισμα ή το ολοκλήρωμα εκτείνεται υπεράνω διαφορετικών συνόλων n-1 τιμών y. Για παραδειγμα, στην ειδική περίπτωση n=2 τυχαίων μεταβλητών X_1, X_2, ισχύει ότι

$$P_Y(y) = \sum_{y_1} P_{X_1,X_2}(y_1, y - y_1) = \sum_{y_2} P_{X_1,X_2}(y - y_2, y_2)$$

και

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{x_1,x_2}(y_1, y - y_1) dy_1 = \int_{-\infty}^{+\infty} f_{x_1,x_2}(y - y_2, y_2) dy_2$$

όπου $Y = X_1 + X_2$.

Παράδειγμα: Έστω X_1, X_2 δύο τυχαίες μεταβλητές με από κοινού συνάρτηση κατανομής

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2(1-\rho^2)} \right],$$

$-\infty < x_1, x_2 < \infty$, $|\rho| \le 1$.

Είναι, δηλαδή, η από κοινού συνάρτηση κατανομής των μεταβλητών αυτών η διμεταβλητή κανονική με παραμέτρους $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$ και ρ. Τότε, κάνοντας χρήση του τελευταίου αποτελέσματος, έχουμε ότι η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $Y = X_1 + X_2$ είναι

$$f_Y(y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp\left[-\frac{y_1^2 - 2\rho y_1 (y - y_1) + (y - y_1)^2}{2(1-\rho^2)} \right] dy_1$$

$$= \frac{\exp\left[-\frac{y^2}{4(1+\rho)} \right]}{\sqrt{2\pi(2)(1+\rho)}} \int_{-\infty}^{+\infty} \frac{\exp\left(-\frac{1}{2} \frac{(y_1 - \frac{1}{2}y)^2}{2(1-\rho)/2} \right)}{\sqrt{2\pi(1-\rho)/2}} dy_1$$

$$= \frac{\exp\left[-\frac{1}{2}\frac{y^2}{2(1+\rho)}\right]}{\sqrt{2\pi(2)(1+\rho)}}, \quad -\infty < y < \infty,$$

η οποία είναι η συνάρτηση πυκνότητας πιθανότητας μιας κανονικής μεταβλητής με μέση τιμή 0 και διασπορά $\sigma^2 = 2(1+\rho)$.

Οπως φαίνεται από το τελευταίο παράδειγμα, η μέθοδος προσδιορισμού της κατανομής του αθροίσματος τυχαίων μεταβλητών που βασίζεται στις τελευταίες σχέσεις, παρέχει την δυνατότητα αντιμετώπισης περιπτώσεων όπου οι τυχαίες μεταβλητές δεν είναι ανεξάρτητες. Στην περίπτωση της ανεξαρτησίας, είναι προφανές ότι οι σχέσεις που δόθηκαν παραπάνω παίρνουν την απλούστερη μορφή:

$$P_Y(y) = \sum_{y_1} P_{X_1}(y_1) P_{X_2}(y - y_1) = \sum_{y_2} P_{X_1}(y - y_2) P_{X_2}(y_2)$$

και

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{x_1}(y_1)\, f_{x_2}(y - y_1)\, dy_1 = \int_{-\infty}^{+\infty} f_{x_1}(y - y_2)\, f_{x_2}(y_2)\, dy_2,$$

αντίστοιχα.

Παρατήρηση: Στην περίπτωση που ρ=0, δηλαδή στην περίπτωση που οι τυχαίες μεταβλητές X_1 και X_2 είναι ανεξάρτητες, η κατανομή της μεταβλητής $Y = X_1 + X_2$ εξακολουθεί να είναι κανονική. Γενικότερα, μπορεί να αποδειχθεί ότι το άθροισμα n ανεξαρτήτων κανονικών μεταβλητών είναι κανονική μεταβλητή με μέση τιμή το άθροισμα των μέσων τιμών και διασπορά το άθροισμα των διασπορών. Η συνάρτηση πιθανότητας (αντίστοιχα πυκνότητας πιθανότητας) της τυχαίας μεταβλητής $Y = X_1 + X_2$, όπως αυτή προσδιορίζεται από την τελευταία μέθοδο που περιγράψαμε (τόσο στην περίπτωση εξαρτημένων μεταβλητών όσο και στην περίπτωση ανεξαρτήτων μεταβλητών), ονομάζεται *συνέλιξη* (convolution) των συναρτήσεων πιθανότητας (αντίστοιχα των συναρτήσεων πυκνότητας πιθανότητας) των τυχαίων μεταβλητών X_1 και X_2.

Ο συμβολισμός που χρησιμοποιείται συνήθως στην διεθνή βιβλιογραφία για την συνέλιξη δύο συναρτήσεων πιθανότητας ή αντίστοιχα δύο συναρτήσεων πυκνότητας πιθανότητας είναι

$$p_Y = p_{X_1} * p_{X_2}$$

και

$$f_Y = f_{X_1} * f_{X_2},$$

αντίστοιχα.

Παράδειγμα: Εστω ότι X_1 και X_2 είναι ανεξάρτητες και ισόνομες διακριτές τυχαίες μεταβλητές που ακολουθούν τη γεωμετρική κατανομή με παράμετρο p. Δηλαδή,

$$p_{X_i} = P_{X_i}(x) = \begin{cases} p(1-p)^x, & x = 0,1,2,... \\ \\ 0, & \text{διαφορετικά} \end{cases}, \quad i = 1,2$$

Τότε, η κατανομή της τυχαίας μεταβλητής $Y = X_1 + X_2$ είναι η συνέλιξη των συναρτήσεων πιθανότητας p_{X_1} και p_{X_2}. Θα έχουμε, δηλαδή, για $y = 0,1,2,...$

$$P_Y(y) = p_{X_1} * p_{X_2} = \sum_{y_1=0}^{y} P_{X_1}(y_1) P_{X_2}(y - y_2)$$

$$= \sum_{y_1=0}^{y} p^2 (1-p)^y$$

$$= (y+1)\, p^2 (1-p)^y.$$

Όπως αναμενόταν, η κατανομή της τυχαίας μεταβλητής Y είναι η αρνητική διωνυμική με παραμέτρους r=2 και p. (Ας σημειωθεί ότι $\begin{bmatrix} y+2-1 \\ 2-1 \end{bmatrix} = y+1$).

Παράδειγμα: Έστω ότι οι τυχαίες μεταβλητές X_1 και X_2 είναι ανεξάρτητες X^2 μεταβλητές με n_1 και n_2 βαθμούς ελευθερίας, αντίστοιχα. Δηλαδή,

$$f_{X_i}(x_i) = \begin{cases} \dfrac{x_i^{\frac{1}{2}n_i - 1} e^{-\frac{1}{2}x_i}}{2^{\frac{1}{2}n_i}\, \Gamma(n_i/2)}, & \text{αν } 0 < x_i < \infty \\ \\ 0, & \text{διαφορετικά} \end{cases}, \quad i = 1,2$$

Τότε, η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής $Y = X_1 + X_2$ (η συνέλιξη των συναρτήσεων πυκνότητας πιθανότητας των τυχαίων μεταβλητών X_1 και X_2) δίνεται από τον τύπο

$$f_Y(y) = \frac{e^{-\frac{1}{2}y}}{2^{\frac{1}{2}(n_1+n_2)}\Gamma(n_1/2)\Gamma(n_2/2)} \int_0^y (y - y_2)^{\frac{1}{2}n_1-1} y_2^{\frac{1}{2}n_2-1} dy_2$$

$$= \frac{y^{\frac{1}{2}(n_1+n_2)-1} e^{-\frac{1}{2}y}}{2^{\frac{1}{2}(n_1+n_2)}\Gamma[(n_1+n_2)/2]} \int_0^1 \frac{z^{\frac{1}{2}n_2-1}(1-z)^{\frac{1}{2}n_1-1}}{B(n_1/2, n_2/2)} dz$$

$$= \frac{y^{\frac{1}{2}(n_1+n_2)-1} e^{-\frac{1}{2}y}}{2^{\frac{1}{2}(n_1+n_2)}\Gamma[(n_1+n_2)/2]}$$

Δηλαδή, η κατανομή της τυχαίας μεταβλητής Y είναι η X^2 με $n_1 + n_2$ βαθμούς ελευθερίας. Το αποτέλεσμα αυτό αναμενόταν, αφού η κατανομή X^2 είναι ειδική περίπτωση της κατανομής Γάμμα και, όπως ήδη έχει αποδειχθεί (βλέπε π.χ. *Ε. Ξεκαλάκη και Ι. Πανάρετο, 1993*), το άθροισμα n ανεξαρτήτων Γάμμα μεταβλητών με κοινή παράμετρο θ, ακολουθεί την κατανομή Γάμμα.

Παρατήρηση: Γενικεύοντας το παραπάνω αποτέλεσμα, ισχύει επομένως ότι αν $X_1, X_2, ..., X_m$ είναι m ανεξάρτητες X^2 μεταβλητές με $n_1, n_2, ..., n_m$ βαθμούς ελευθερίας αντίστοιχα ($X_i \sim X^2_{n_i}$, i=1,2,...,m), τότε,

$X_1 + X_2 + ... + X_m \sim X^2_{n_1+n_2+...+n_m}$.

Το αποτέλεσμα αυτό μπορεί να ληφθεί είτε ως άμεση συνέπεια του αντίστοιχου αποτελέσματος για ανεξάρτητες Γάμμα τυχαίες μεταβλητές, είτε ως αποτέλεσμα πολλαπλών διαδοχικών εφαρμογών της διαδικασίας της συνέλιξης. Πράγματι, παρατηρώντας ότι

$$X_1 + X_2 + ... + X_m = [...[(X_1 + X_2) + X_3] + ... + X_{m-1}] + X_m,$$

θα έχουμε, σύμφωνα με το τελευταίο παράδειγμα, διαδοχικά ότι

$$(X_1 + X_2) \sim X^2_{n_1+n_2},$$

$$(X_1 + X_2) + X_3 \sim X^2_{n_1+n_2+n_3},$$

. .

. .

. .

$$X_1 + X_2 + \ldots + X_m \sim X^2_{n_1+n_2+\ldots+n_m},$$

Είναι, δηλαδή, στην περίπτωση αυτή η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής Y η συνέλιξη των συναρτήσεων πυκνότητας πιθανότητας των μεταβλητών X_i , i=1,2,...,m. Συνελίξεις περισσοτέρων από δύο συναρτήσεων πυκνότητας πιθανότητας (αντίστοιχα συναρτήσεων πιθανότητας) ονομάζονται *πολλαπλές συνελίξεις* και συμβολίζονται με $f_Y = f_{X_1} * f_{X_2} * \ldots * f_{X_n}$, για την συνεχή περίπτωση και με $p_Y = p_{X_1} * p_{X_2} * \ldots * p_{X_n}$, για την διακριτή περίπτωση.

Στην περίπτωση που οι τυχαίες μεταβλητές X_i, i=1,2,...,n είναι ανεξάρτητες, η συνέλιξη των n περιθωρίων συναρτήσεων πυκνότητας πιθανότητας, για την συνεχή περίπτωση, ή των n περιθωρίων συναρτήσεων πιθανότητας, για την διακριτή περίπτωση, ορίζονται αντίστοιχα ως εξής:

$$f_Y(y) = (f_{X_1} * f_{X_2} * \ldots * f_{X_n})(y)$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty} f_{X_1}(y_1) f_{X_2}(y_2 - y_1)\ldots f_{X_n}(y - y_{n-1})dy_1 dy_2 \ldots dy_{n-1}$$

και

$$p_Y = p_Y(y) = (p_{X_1} * p_{X_2} * \ldots * p_{X_n})(y)$$

$$= \sum_{y_{n-1}}\sum_{y_{n-2}}\ldots\sum_{y_1} P_{X_1}(y_1) P_{X_2}(y_2 - y_1)\ldots P_{X_n}(y - y_{n-1}).$$

Σημείωση: Αξίζει να σημειωθεί ότι η συνέλιξη έχει πολλές από τις ιδιότητες των μαθηματικών πράξεων, όπως την *αντιμεταθετικότητα* και την *προσεταιριστικότητα*. Ισχύει δηλαδή ότι

$$p_{X_1} * p_{X_2} = p_{X_2} * p_{X_2} \qquad \text{(αντιμεταθετικότητα)}$$

και

$$(p_{X_1} * p_{X_2}) * p_{X_3} = p_{X_2} * (p_{X_1} * p_{X_3}) \qquad \text{(προσεταιριστικότητα)},$$

στην διακριτή περίπτωση, με αντίστοιχους ορισμούς για την συνεχή περίπτωση. Οι συνελίξεις έχουν επίσης μονάδα, δηλαδή, αν X_0 είναι μία τυχαία μεταβλητή με συνάρτηση πιθανότητας

$$p_0 \equiv p_{X_0}(x) = \begin{cases} 1, & \text{αν } x = 0 \\ 0, & \text{διαφορετικά} \end{cases}$$

τότε, για οποιαδήποτε τυχαία μεταβλητή X με συνάρτηση πιθανότητας
$p = P_X(x)$ ισχύει ότι $p_X * p_0 = p_X$.
Συνοψίζοντας τα αποτελέσματα του παρόντος κεφαλαίου, και λαβαίνοντας υπόψη τις ιδιότητες της μέσης τιμής και της διασποράς, απαριθμούμε τις προκύπτουσες ιδιότητες διαφόρων κατανομών, οι οποίες εφαρμόζονται πολύ συχνά σε σχέση με πρακτικά προβλήματα:
(1) Κανονική Κατανομή: Αν $X_1, X_2,..., X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim N(\mu_i, \sigma_i^2)$, i=1,2,...,n, τότε

$$Y = \sum_{i=1}^{n} \alpha_i X_i \sim N\left(\sum_{i=1}^{n} \alpha_i \mu_i , \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2 \right).$$

(2) Διμεταβλητή κανονική κατανομή: Αν $X_1, X_2,..., X_n$ είναι τυχαίες μεταβλητές με από κοινού κατανομή την διμεταβλητή κανονική κατανομή, δηλαδή $(X_1, X_2) \sim$ διμεταβλητή $N(\mu_1, \mu_2 ; \sigma_1^2, \sigma_2^2 ; \rho)$,τότε

$$\alpha_1 X_1 + \alpha_2 X_2 + c \sim N(\alpha_1 \mu_1 + \alpha_2 \mu_2 + c , \alpha_1 \sigma_1^2 + \alpha_2 \sigma_2^2 + 2\alpha_1 \alpha_2 \rho) .$$

(3) Διωνυμική κατανομή: Αν $X_1, X_2,..., X_m$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim b(x; n_i , p)$, i=1,2,...,m, τότε

$$Y = \sum_{i=1}^{m} X_i \sim b\left(y; \sum_{i=1}^{m} n_i , p \right).$$

(4) Κατανομή Poisson: Αν $X_1, X_2,..., X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim P(x; \lambda_i)$, i=1,2,...,n, τότε

$$Y = \sum_{i=1}^{n} X_i \sim P\left(y ; \sum_{i=1}^{n} \lambda_i \right).$$

(5) Αρνητική διωνυμική κατανομή: Αν $X_1, X_2,..., X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim NB(x; r_i , p)$, i=1,2,...,n, τότε

$$Y = \sum_{i=1}^{n} X_i \sim NB\left(y ; \sum_{i=1}^{n} r_i , p \right).$$

(6) Κατανομή Γάμμα: Αν $X_1, X_2 ,..., X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές με $X_i \sim \gamma(x; \alpha_i , \theta)$, i=1,2,...,n, τότε

$$Y = \sum_{i=1}^{n} X_i \sim \gamma\left(y; \sum_{i=1}^{n} \alpha_i , \theta \right).$$

(7) Κατανομή X^2 : Αν $X_1, X_2 ,..., X_k$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim X^2_{n_i}$, i=1,2,...,k, τότε

$$Y = \sum_{i=1}^{k} X_i \sim X^2_{n_i + n_2 + ... + n_k}.$$

(8) Εκθετική κατανομή: Αν $X_1, X_2 ,...,X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim \exp(\theta)$, i=1,2,...,n, τότε

$$Y = \sum_{i=1}^{n} X_i \sim \gamma\left(y; n ; \theta\right).$$

(9) Γεωμετρική κατανομή : Αν $X_1, X_2,...,X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές και $X_i \sim G(x; p)$, τότε

$$Y = \sum_{i=1}^{n} X_i \sim NB\left(y ; n; p\right).$$