ΚΕΦΑΛΑΙΟ 8

ΜΕΡΙΚΕΣ ΕΙΔΙΚΕΣ ΣΥΝΕΧΕΙΣ ΚΑΤΑΝΟΜΕΣ

Στις ενότητες που ακολουθούν εξετάζουμε συνεχείς κατανομές με ευρεία χρήση στις εφαρμογές. Σε αυτές περιλαμβάνονται η ομοιόμορφη, η εκθετική, η Γάμμα και η Βήτα. Η σπουδαιότερη όμως συνεχής κατανομή, κυρίως λόγω των πολλών εφαρμογών της, είναι η κανονική. Λόγω της σπουδαιότητάς της, αποτελεί ξεχωριστό κεφάλαιο το οποίο ακολουθεί.

Μια άλλη ειδική κατηγορία συνεχών κατανομών με πλήθος εφαρμογών στην Στατιστική αποτελούν οι λεγόμενες δειγματικές κατανομές *(sampling distributions).* Τέτοιες κατανομές είναι η t, η X^2 και η F. Επειδή οι κατανομές αυτές (στις οποίες περιλαμβάνεται και η κανονική) προκύπτουν κυρίως σε σχέση με την Στατιστική, εξετάζονται επίσης σε χωριστό κεφάλαιο, παρότι είναι και αυτές συνεχείς κατανομές.

Η ΟΜΟΙΟΜΟΡΦΗ ΚΑΤΑΝΟΜΗ

Ορισμός: Έστω Χ μια (απόλυτα) συνεχής τυχαία μεταβλητή ορισμένη στο διάστημα [a, β] με συνάρτηση πυκνότητας πιθανότητας.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ \\ 0 & \text{διαφορετικά} \end{cases}$$

Στην περίπτωση αυτή, λέμε ότι το Χ ακολουθεί την *συνεχή ομοιόμορφη κατανομή (continuous uniform distribution) με παραμέτρους* a *και* b και συμβολίζουμε X~U (a,b). Είναι προφανές ότι η ομοιόμορφη κατανομή, όπως ορίσθηκε, είναι μια καλά ορισμένη κατανομή.

Μοντέλα που οδηγούν στην ομοιόμορφη κατανομή

Η ομοιόμορφη κατανομή θα μπορούσε να θεωρηθεί ως προκύπτουσα στο πλαίσιο τυχαίου πειράματος κατά το οποίο επιλέγεται τυχαία ένα σημείο από το διάστημα [a,b] με "δίκαιο" τρόπο.

Πρόταση: Αν $X \sim U(a,b)$, τότε

$$E(X) = \frac{b+a}{2} \quad \text{και} \quad \Delta(X) = \frac{(a-b)^2}{12}$$

Απόδειξη:

$$E(X) = \int_a^b xf(x)dx = \frac{1}{b-a}\left.\frac{x^2}{2}\right|_a^b = \frac{b+a}{2}$$

$$E(X^2) = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3-a^3}{3(b-a)}$$

και

$$\Delta(X) = E(X^2) - \{E(X)\}^2 = \frac{b^3-a^3}{3(b-a)} - \frac{(b+a)^2}{4}$$

$$= \frac{4(a^2+ab+b^2)-3(b^2+a^2+2ab)}{12} = \frac{(a-b)^2}{12}$$

Παρατήρηση: Ισχύει ότι

$$F(x) = \int_{-\infty}^x f(t)dt = \frac{x-a}{b-a}, \quad a \le x \le b$$

Η περισσότερο χρησιμοποιούμενη ομοιόμορφη κατανομή είναι η U(0,1). Η κατανομή αυτή χρησιμοποιείται για την τεχνητή κατασκευή (προσομοίωση) τυχαίων δειγμάτων από διακριτές και συνεχείς κατανομές.

Παράδειγμα: Η ζήτηση για κάποιο προϊόν ακολουθεί την ομοιόμορφη κατανομή στο διάστημα (9500, 10500), σε πακέτα ανά μήνα. Για να αναλυθούν ευκολότερα τα δεδομένα αυτά, εφαρμόζουμε ένα γραμμικό μετασχηματισμό έτσι ώστε τα μετασχηματισμένα δεδομένα να έχουν τυποποιημένη μορφή (μέση τιμή 0 και διασπορά 1). Να βρεθεί η συνάρτηση πυκνότητας πιθανότητας των μετασχηματισμένων δεδομένων.

Λύση: Έστω X η ζήτηση για το προϊόν. Έχουμε ότι

$$X \sim U(9500,\ 10500)$$

Επομένως,

$$E(X) = \frac{9500+10500}{2} = 10000 \quad \text{και}$$

$$\Delta(X) = \frac{(10500-9500)^2}{12} = \frac{(1000)^2}{12}$$

Σύμφωνα με τον ορισμό της τυποποίησης ο απαιτούμενος μετασχηματισμός είναι

$$Y = \frac{X-10000}{(1000/\sqrt{12})}, \qquad -\sqrt{3} \le Y \le \sqrt{3}$$

όπου Y είναι τα τυποποιημένα δεδομένα.
Η συνάρτηση κατανομής της Y είναι

$$F(y) = P(Y \le y) = P\left(\frac{X-10000}{(1000/\sqrt{12})} \le y\right) = P\left(X \le y\frac{1000}{\sqrt{12}} + 10000\right)$$

$$= \int_{9500}^{y\frac{1000}{\sqrt{12}}+1000} \frac{1}{1000}\,dx$$

$$= \frac{1}{1000}\left(y\frac{1000}{\sqrt{12}} + 10000 - 9500\right) = \frac{y}{\sqrt{12}} - \frac{1}{2}$$

και παραγωγίζοντας βρίσκουμε ότι

$$f(y) = F'(y) = \frac{1}{\sqrt{12}}$$

δηλαδή τελικά η τυποποιημένη Y ακολουθεί επίσης την ομοιόμορφη κατανομή στο διάστημα $[-\sqrt{3}, \sqrt{3}]$. Είναι εύκολο να επιβεβαιωθεί ότι

$$E(Y) = 0 \quad \text{και} \quad \Delta(X) = 1.$$

Η ΕΚΘΕΤΙΚΗ ΚΑΤΑΝΟΜΗ

Ορισμός: Μια συνεχής τυχαία μεταβλητή X με συνάρτηση κατανομής

$$F(x)=1-e^{-x/\theta}, \qquad x > 0, \theta > 0$$

λέγεται ότι ακολουθεί την *εκθετική κατανομή (exponential distribution) με παράμετρο θ.*
Συμβολικά X~exp(θ).

Η συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής δίνεται από την σχέση

$$f(x) = F'(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 \le x < \infty$$

Ο υπολογισμός πιθανοτήτων της εκθετικής κατανομής γίνεται ή με ολοκλήρωση ή με χρήση της συνάρτησης κατανομής και των πινάκων της κατανομής Poisson. Για παράδειγμα, αν X~exp(θ=10), τότε

$$P(X < 20) = F(20) = 1-e^{-20/10} = 1-e^{-2} = 1-(0.135)$$

Εναλλακτικά

$$P(X < 20) = \int_0^{20} \frac{1}{20} e^{-x/20} dx = 1 - e^{-2}$$

Επίσης,

$$P(X > 15) = 1-P(X \le 15) = 1-(1-e^{-15/10}) = e^{-1.5} = 0.233$$

Ιδιότητες: Αν X~exp(θ), τότε
1) $E(X) = \theta$, $\Delta(X) = \theta^2$
2) $P(X > \theta) < ½$

Μοντέλα που οδηγούν στην εκθετική κατανομή

Θεώρημα: Ο χρόνος αναμονής μέχρις ότου πραγματοποιηθεί το πρώτο "γεγονός" σε μια διαδικασία Poisson με παράμετρο λ ακολουθεί την εκθετική κατανομή με παράμετρο $\frac{1}{\lambda}$.

Απόδειξη: Έστω W ο χρόνος αναμονής μέχρι το πρώτο γεγονός (W≥0). F(w)=P(W≤w)=1-P(W>w)

=1-P(0 γεγονότα στο διάστημα [0, w])=1-$e^{-\lambda w}$

Γενικότερα, ισχύει ότι ο χρόνος που μεσολαβεί μεταξύ δύο διαδοχικών "γεγονότων" μιας διαδικασίας Poisson ακολουθεί την εκθετική κατανομή. Έτσι, για παράδειγμα, ο χρόνος που μεσολαβεί μεταξύ δύο διαδοχικών ισχυρών σεισμών, ο χρόνος μεταξύ δύο διαδοχικών κλήσεων σε ένα τηλεφωνικό κέντρο, ο χρόνος μεταξύ δύο διαδοχικών ατυχημάτων σε μια διασταύρωση κ.λ.π. ακολουθεί την εκθετική κατανομή.

Παράδειγμα: Έστω ότι ο αριθμός των πελατών που επισκέπτονται μια τράπεζα ακολουθεί την κατανομή Poisson με μέσο ρυθμό 6 πελάτες την ώρα. α) Να υπολογισθεί η πιθανότητα ώστε ο χρόνος που μεσολαβεί μεταξύ δύο αφίξεων να υπερβαίνει τα 15 λεπτά. β) Να υπολογισθεί ο μέσος χρόνος που μεσολαβεί μεταξύ διαδοχικών αφίξεων.

Λύση: α) Έστω Χ ο χρόνος αναμονής. Επειδή ο αριθμός των αφίξεων μπορεί να προσεγγισθεί με την κατανομή Poisson με παράμετρο (μέσο αριθμών αφίξεων ανά λεπτό)

$$\lambda=\frac{6}{60}=\frac{1}{10}, \text{ έχουμε ότι } X\sim\exp(=10)$$

Θα έχουμε

$$P(X > 15) = e^{-15/10} = e^{-1.5} = 0.223$$

β) Ε(Χ) = θ = 10 άτομα/λεπτό.

Πρόταση: (έλλειψη μνήμης της εκθετικής κατανομής)

Αν X~exp(θ) τότε

$$P(X > x+y \mid X > x) = P(X > y)$$

(Η δεσμευμένη πιθανότητα ο χρόνος αναμονής X να υπερβεί το x+y δοθέντος ότι έχει ήδη υπερβεί το x ισούται με την πιθανότητα το X να υπερβεί το y. Ο χρόνος δηλαδή που έχει ήδη διαρρεύσει δεν επηρεάζει τον χρόνο που απαιτείται για την πραγματοποίηση ενός "γεγονότος" σε μια διαδικασία Poisson).

Απόδειξη: Από τον ορισμό της δεσμευμένης πιθανότητας έχουμε ότι

$$P(X > x + y \mid X > x) = \frac{P(X > x + y, X > x)}{P(X > x)} = \frac{P(X > x + y)}{P(X > x)}$$

Το συμπέρασμα του θεωρήματος είναι άμεση συνέπεια του γεγονότος ότι $P(X > z) = e^{-z/\theta}$.

Παράδειγμα: Στο προηγούμενο παράδειγμα να υπολογισθεί η πιθανότητα ώστε ο χρόνος που μεσολαβεί ανάμεσα σε δύο διαδοχικές αφίξεις πελατών να υπερβεί τα 25 λεπτά δοθέντος ότι έχουν ήδη περάσει 10 λεπτά χωρίς άφιξη.

Λύση:

$$P(X > 25 \mid X > 10) = P(X > 15) = 0.223$$

Παρατήρηση: Μια σύγκριση ανάμεσα στην εκθετική και την γεωμετρική κατανομή οδηγεί στο συμπέρασμα ότι η δεύτερη είναι το συνεχές ανάλογο της πρώτης. Πράγματι, η εκθετική κατανομή αναφέρεται στον χρόνο που μεσολαβεί μεταξύ δύο διαδοχικών "γεγονότων" μιας διαδικασίας Poisson ενώ η γεωμετρική κατανομή αναφέρεται στον αριθμό των αποτυχιών μεταξύ δύο επιτυχιών σε μια ακολουθία δοκιμών Bernoulli. Η αντιστοιχία δηλαδή είναι

[χρόνος αναμονής]	⟷	[αριθμός αποτυχιών]
["γεγονότα" Poisson]	⟷	["επιτυχίες" σε μια ακολουθία δοκιμών Bernoulli]
[εκθετική]	⟷	[γεωμετρική]

Σημείωση: Η ιδιότητα της “έλλειψης μνήμης” της εκθετικής κατανομής που αναφέρθηκε στην προηγούμενη πρόταση υποδηλώνει και μια άλλη εφαρμογή της εκθετικής κατανομής. Συγκεκριμένα, δίνει την δυνατότητα χρησιμοποίησης της εκθετικής κατανομής ως κατανομής του χρόνου ζωής όταν δεν υπάρχει φθορά λόγω χρόνου. Για παράδειγμα, έστω X ο χρόνος ζωής (σε ώρες) μιας λυχνίας τηλεόρασης. Τότε $P(X>c)$ εκφράζει την πιθανότητα ότι μια καινούργια λυχνία θα διαρκέσει τουλάχιστον c ώρες ενώ $P(X>b+c|X>b)$ είναι η πιθανότητα ώστε μια μεταχειρισμένη λυχνία που έχει ήδη χρησιμοποιηθεί για b ώρες να διαρκέσει για c επιπλέον ώρες. Εάν ο χρόνος ζωής των λυχνιών αυτών ακολουθεί την εκθετική κατανομή, η ιδιότητα της “έλλειψης μνήμης” της εκθετικής κατανομής συνεπάγεται ότι οι δύο αυτές πιθανότητες είναι ίσες για οποισδήποτε τιμές των b και c. Η πιθανότητα βλάβης στις επόμενες c ώρες δεν εξαρτάται από τον χρόνο που η λυχνία βρίσκεται ήδη σε χρήση. Ο χρόνος λειτουργίας δηλαδή δεν επιφέρει φθορά. Οι μεταχειρισμένες λυχνίες επομένως στην περίπτωση αυτή είναι εξ’ ίσου καλές με τις αμεταχείριστες.

Παράδειγμα: Έστω ότι οι αστέρες είναι τυχαία κατανεμημένοι στον χώρο. Να βρεθεί η κατανομή πιθανότητας της απόστασης ενός τυχαίου σημείου στον χώρο από τον πλησιέστερο αστέρα.

Λύση: Κάτω από την υπόθεση της τυχαίας κατανομής των αστέρων στον χώρο έχουμε ότι η πιθανότητα να βρεθούν x αστέρια σε όγκο V του χώρου είναι

$$P(X = x) = e^{-\lambda V} \frac{(\lambda V)^x}{x!}, \qquad x=0, 1, 2, \dots$$

όπου λ είναι ο μέσος αριθμός αστέρων στην μονάδα του όγκου. Έστω R η απόσταση ενός τυχαίου σημείου Σ στον χώρο από τον πλησιέστερο αστέρα. Τότε $R \le r$ τότε και μόνο τότε αν υπάρχει τουλάχιστον ένας αστέρας στην σφαίρα με κέντρο το Σ και ακτίνα r. Ο όγκος της σφαίρας είναι $\frac{4}{3}\pi r^3$ και επομένως

$$P(R \le r) = 1 - e^{-4\lambda\pi r^3/3}, \qquad r > 0$$

(Αυτό γιατί η πιθανότητα να υπάρχει ένας τουλάχιστον αστέρας σε όγκο V είναι 1-P(X=0)=1-$e^{-\lambda V}$).
Παραγωγίζοντας ως προς r έχουμε την συνάρτηση πυκνότητας πιθανότητας του R.
Είναι ενδιαφέρον να σημειωθεί ότι αν θεωρήσουμε σαν Σ την θέση ενός αστέρα τότε το R συμβολίζει την απόστασή του από τον πλησιέστερο προς αυτόν αστέρα.

Η ΚΑΤΑΝΟΜΗ ΓΑΜΜΑ

Πριν προχωρήσουμε στον ορισμό της κατανομής γάμμα είναι απαραίτητο να ορίσουμε την συνάρτηση γάμμα.

Ορισμός: Η γάμμα συνάρτηση ορίζεται από την σχέση

$$\Gamma(t) = \int_0^\infty e^{-y} y^{t-1} dy, \qquad t>0$$

Επομένως,

$$\Gamma(1) = \int_0^\infty e^{-y} dy = 1, \qquad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Για t>1

$$\Gamma(t) = (t-1)\Gamma(t-1)$$

και επομένως όταν t=n (ακέραιος)

$$\Gamma(n)=(n-1)\Gamma(n-1)=(n-1)(n-2)\gamma(n-2)=(n-1)(n-2)\ldots 1\Gamma(1)\Rightarrow$$
$$\Gamma(n)=(n-1)!$$

Ορισμός: Θα λέμε ότι η (απόλυτα) *συνεχής τυχαία μεταβλητή X ακολουθεί την κατανομή Γάμμα με παραμέτρους* (α, θ) και θα συμβολίζουμε με X ~ γ(x; α, θ), αν

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} e^{-x/\theta} x^{\alpha-1} = \frac{1}{\Gamma^{\alpha}(\alpha)} e^{-x/\theta}\left(\frac{x}{\theta}\right)^{\alpha-1}, \qquad 0\le x<\infty,\ \alpha>0,\ \theta>0$$

Παρατήρηση: Η κανονική γάμμα είναι μια καλά ορισμένη κατανομή αφού

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{0}^{+\infty} \frac{1}{\Gamma^{\alpha}(\alpha)} e^{-x/\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} dx$$

$$= \frac{1}{\Gamma(\alpha)} \int_{-\infty}^{+\infty} e^{-x/\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} d\left(\frac{x}{\theta}\right)$$

$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$$

Σημείωση: Αν X ακολουθεί την κατανομή γάμμα με παραμέτρους α και θ, τότε

$$E(X)=\alpha, \quad \Delta(X)=\alpha\theta^2$$

Σημείωση: Η κατανομή γάμμα με παραμέτρους α και θ=1 λέγεται *τυποποιημένη κατανομή γάμμα*. Δηλαδή, για την τυποποιημένη κατανομή γάμμα

$$f(x) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1}, \quad x\geq 0,\ \alpha>0$$

Σημείωση: Η παράμετρος θ της κατανομής γάμμα ονομάζεται και *παράμετρος κλίμακας (scale parameter)* γιατί τιμές της διάφορες του 1 εκτείνουν ή συμπτύσσουν την συνάρτηση πυκνότητας πιθανότητας κατά την διεύθυνση του άξονα των x.

Υπολογισμός των Πιθανοτήτων της Κατανομής Γάμμα

Όταν η τυχαία μεταβλητή X ακολουθεί την τυποποιημένη κατανομή γάμμα, η συνάρτηση κατανομής του X δίνεται από την σχέση

$$\Phi(x) = P(X \leq x) = \int_{0}^{x} \frac{1}{\Gamma(\alpha)} e^{-y} y^{\alpha-1} dy, \quad x, y\geq 0$$

Η συνάρτηση αυτή ονομάζεται και ελλιπής συνάρτηση γάμμα για προφανείς λόγους. (Πολλές φορές ως *ελλιπής συνάρτηση γάμμα* ορίζεται η παραπάνω συνάρτηση χωρίς το Γ(α)). Για τον υπολογισμό της ελλιπούς συνάρτησης γάμμα υπάρχουν αναλυτικοί πίνακες. Μέρος των πινάκων αυτών ακολουθεί.

Πίνακας

Η ελλιπής συνάρτηση Γάμμα $\Phi(x) = \int_0^x \frac{1}{\Gamma(\alpha)} e^{-y} y^{\alpha-1} dy$

x \ α	1	2	3	4	5	6	7	8	9	10
1	.632	.264	.080	.019	.004	.001	.000	.000	.000	.000
2	.865	.594	.323	.143	.053	.017	.005	.001	.000	.000
3	.950	.801	.577	.353	.185	.084	.034	.012	.004	.001
4	.982	.908	.762	.567	.371	.215	.131	.051	.021	.008
5	.993	.960	.875	.735	.560	.384	.238	.133	.068	.032
6	.998	.983	.938	.849	.715	.554	.398	.256	.153	.084
7	.999	.993	.970	.918	.827	.699	.550	.401	.271	.170
8	1.000	.997	.986	.958	.900	.809	.687	.547	.407	.283
9		.999	.994	.979	.945	.884	.793	.676	.544	.413
10		1.000	.997	.990	.971	.933	.870	.780	.667	.542
11			.999	.995	.985	.962	.921	.857	.768	.659
12			1.000	.998	.992	.980	.954	.911	.845	.758
13				.999	.996	.989	.974	.946	.900	.834
14				1.000	.998	.994	.986	.968	.938	.891
15					.999	.997	.992	.982	.963	.930

Παράδειγμα: Έστω ότι ο χρόνος αντίδρασης Χ ενός ατόμου σε κάποιο διεγερτικό ακολουθεί την τυποποιημένη κατανομή γάμμα με α=2 sec. Να υπολογισθεί:

(α) Η πιθανότητα ώστε ο χρόνος αντίδρασης να είναι περισσότερος από 3 sec και λιγότερος από 5 sec.

(β) Η πιθανότητα ο χρόνος αντίδρασης να είναι περισσότερος από 4 sec.

Λύση:

(α) $P(3<X<5) = F(5) - F(2) = 0.960 - 0.801 = 0.159$

(β) $P(X>4) = 1- p(X\leq 4) = 1-F(4) = 1 - 0.908 = 0.092$

Σημείωση: Η ελλιπής συνάρτηση γάμμα μπορεί να χρησιμοποιηθεί και για τον υπολογισμό πιθανοτήτων μη τυποποιημένων κατανομών γάμμα. Αυτό γιατί ισχύει η εξής πρόταση.

Πρόταση: Έστω ότι η τυχαία μεταβλητή Χ ακολουθεί την κατανομή γάμμα με παραμέτρους α και θ. Τότε για κάθε $x>0$

$$P(X \leq x) = \Phi\left(\frac{x}{\theta}\right)$$

όπου Φ είναι η συνάρτηση κατανομής της τυποποιημένης κατανομής γάμμα.

Απόδειξη: Προφανής.

Παράδειγμα: Έστω ότι ο χρόνος επιβίωσης X (σε εβδομάδες) ενός τυχαία επιλεγμένου ποντικού που εκτίθεται σε ακτινοβολία 240 rads ακολουθεί την κατανομή γάμμα με παραμέτρους α=8 και θ=15. Να υπολογισθούν

(α) Ο αναμενόμενος χρόνος επιβίωσης.

(β) Η τυπική απόκλιση του χρόνου επιβίωσης.

(γ) Η πιθανότητα ότι ο ποντικός θα επιβιώσει για διάστημα μεγαλύτερο των 60 και μικρότερο των 120 εβδομάδων.

Λύση:

(α) $E(X) = \alpha = 120$ εδομάδες

(β) $\Delta(X) = \alpha^2 = 1800$

$$\sigma = \sqrt{\Delta(X)} = \sqrt{1800 = 42.43} \text{ εβδομάδες.}$$

(γ) $P(60<X<120) = P(X<120) - P(X<60) =$

$$= \Phi\left(\frac{120}{15}\right) - \Phi\left(\frac{60}{15}\right) = \Phi(8) - \Phi(4) = 0.547 - 0.051 = 0.496.$$

Μοντέλα που οδηγούν στην Κατανομή Γάμμα

Πρόταση: Έστω X ο χρόνος αναμονής σε μια ανέλιξη Poisson με παράμετρο λ μέχρις ότου συμβεί το n-οστό "γεγονός". Τότε η X ακολουθεί την κατανομή γάμμα με παραμέτρους

$$\alpha=n \text{ και } \theta=1/\lambda$$

Απόδειξη: Έστω Y ο αριθμός των γεγονότων στο διάστημα [0, x]. Τότε ισχύει ότι

$$Y \sim \text{Poisson}(\lambda x)$$

Άρα

$F(x)=P(X\leq x)=1-P(X>x)=$

=1-P[λιγότερα από n γεγονότα στο διάστημα [0,x]]=

$$=1-\sum_{k=0}^{n-1} e^{-\lambda x}\frac{(\lambda x)^k}{k!} = 1-e^{-\lambda x}\sum_{k=1}^{n-1} e^{-\lambda x}\frac{(\lambda x)^k}{k!}$$

Επομένως,

$$f(x)=F'(x)=\lambda e^{-\lambda x}-\sum_{k=1}^{n-1}\left[(-\lambda)e^{-\lambda x}\frac{(\lambda x)^k}{k!}+e^{-\lambda x}\frac{k\lambda(\lambda x)^{k-1}}{k!}\right]$$

$$=\lambda e^{-\lambda x}\left[1+\sum_{k=1}^{n-1}\frac{(\lambda x)}{k!}-\sum_{k=1}^{n-1}\frac{(\lambda x)^{k-1}}{(k-1)!}\right]$$

$$=\lambda e^{-\lambda x}\left[1+\sum_{k=1}^{n-1}\frac{(\lambda x)^k}{k!}-\sum_{k=0}^{n-2}\frac{(\lambda x)^k}{k!}\right]$$

$$=\lambda e^{-\lambda x}\left[1+\sum_{k=1}^{n-2}\frac{(\lambda x)^k}{k!}+\frac{(\lambda x)^{k-1}}{(k-1)!}-1-\sum_{k=1}^{n-2}\frac{(\lambda x)^k}{k!}\right]$$

$$=\frac{\lambda}{\Gamma(n)}e^{-\lambda x}(\lambda x)^{n-1}$$

δηλαδή, X~Γάμμα (α=n, θ=1/λ).

Σημείωση: Όταν η παράμετρος α της κατανομής γάμμα είναι ακέραιος αριθμός (α=n) η κατανομή γάμμα λέγεται και *κατανομή n-Erlang.*

Παρατήρηση: Είναι προφανές ότι για α=1 η κατανομή γάμμα συμπίπτει με την εκθετική κατανομή. Αυτό άλλωστε είναι φυσικό λόγω της δυνατότητας ερμηνείας των δύο κατανομών ως κατανομών του χρόνου αναμονής σε μια ανέλιξη Poisson.

Πρόταση: Έστω X_1, X_2, …, X_n ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με παράμετρο θ. Τότε η τυχαία μεταβλητή

$$X= X_1 + X_2 + \ldots + X_n$$

ακολουθεί την κατανομή γάμμα με παραμέτρους n και θ.

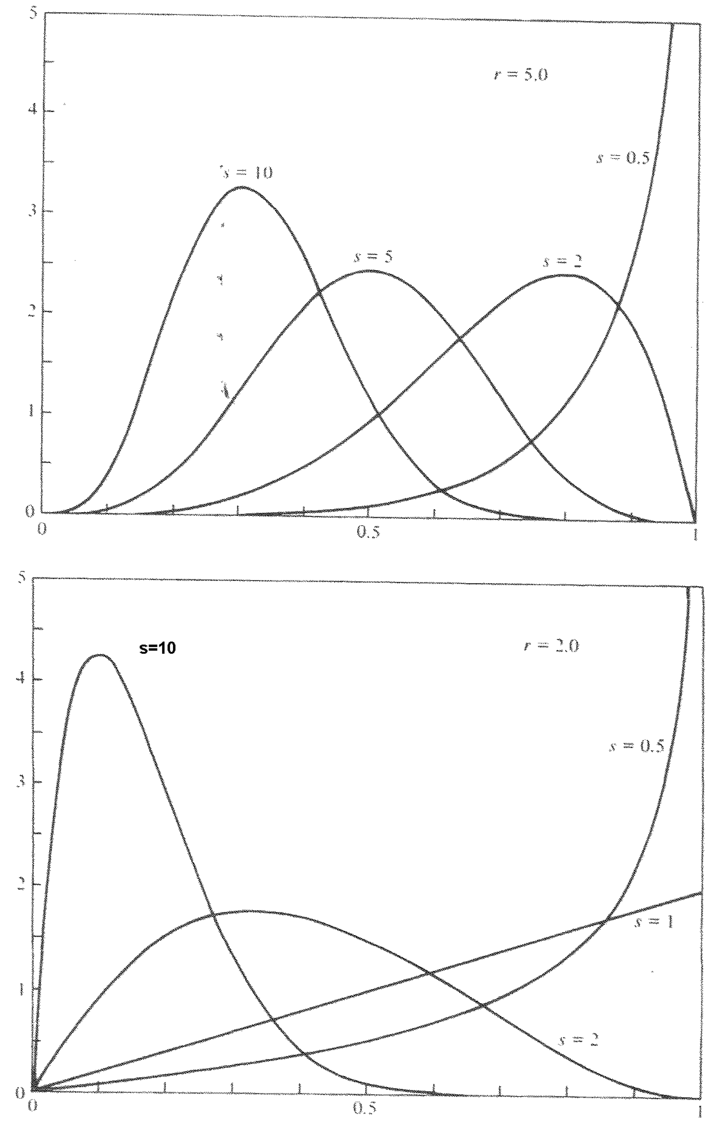
Απόδειξη: Η πρόταση αυτή αποδεικνύεται από το γεγονός ότι η τυχαία μεταβλητή X_i μπορεί να θεωρηθεί ως ο χρόνος αναμονής μεταξύ του (i-1) και i "γεγονότος" σε μια ανέλιξη Poisson. Έχει ήδη αποδειχθεί ότι τα X_i ακολουθούν την εκθετική κατανομή.
$X= X_1 + X_2 + \ldots + X_n$ είναι ο συνολικός χρόνος αναμονής έως ότου συμβεί το n-οστό γεγονός στην ανέλιξη Poisson. Έχει δε βρεθεί ότι το X ακολουθεί την κατανομή γάμμα.

Με τον ίδιο τρόπο μπορεί να αποδειχθεί ότι αν $X_1, X_2, \ldots, X_n$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την κατανομή γάμμα με παραμέτρους α_i και θ, i=1, 2, …, n, τότε η τυχαία μεταβλητή $X_1 + X_2 + \ldots + X_n$ ακολουθεί την κατανομή γάμμα με παραμέτρους $\sum_{i=1}^{n} \alpha_i$ και θ.

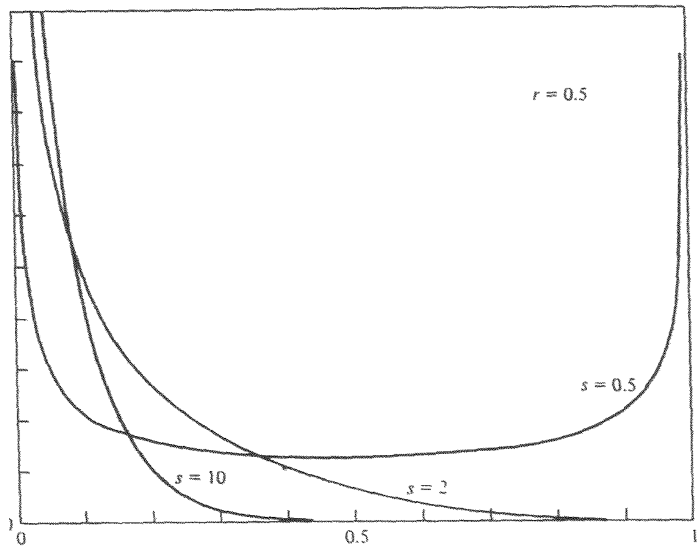
Σημείωση: Η σύνδεση της κατανομής γάμμα με την κατανομή του χρόνου αναμονής σε μία ανέλιξη Poisson οδηγεί στο συμπέρασμα ότι η κατανομή γάμμα είναι το συνεχές ανάλογο της αρνητικής διωνυμικής κατανομής (όπως η εκθετική κατανομή είναι το συνεχές ανάλογο της γεωμετρικής κατανομής). Εδώ ο χρόνος αναμονής (κατανομή γάμμα) αναφέρεται στην πραγματοποίηση του n-οστού γεγονότος και ο αριθμός των αποτυχιών (αρνητική διωνυμική κατανομή) στις αποτυχίες μέχρι την n-οστή επιτυχία.

Η ΚΑΤΑΝΟΜΗ ΒΗΤΑ *(Beta Distribution)*

Σε σχέση με πολλές πρακτικές εφαρμογές, χρησιμοποιούμε απόλυτα συνεχείς τυχαίες μεταβλητές, των οποίων το εύρος των τιμών είναι το διάστημα (0,1). Συχνά, η ακριβής μορφή της συνάρτησης πυκνότητας πιθανότητας δεν είναι γνωστή, αλλά υπάρχουν ορισμένες ενδείξεις ότι η πυκνότητα παρουσιάζει μέγιστο κοντά σε κάποια τιμή. Η κατανομή Βήτα, η οποία έχει δύο θετικές παραμέτρους, έχει συνάρτηση πυκνότητας πιθανότητας η οποία μπορεί να έχει μία ποικιλία μορφών ανάλογα με την επιλογή των τιμών των παραμέτρων, όπως φαίνεται στο σχήμα που ακολουθεί.



Σχήμα: Γραφήματα της συνάρτησης πυκνότητας της κατανομής βήτα για (α) r=5.0, s=0.5, 2.0, 5.0, 10.0; (β) r=2.0, s=0.5,. 1.0, 2.0, 10.0; (συνεχίζεται)



Σχήμα (συνέχεια): Γραφήματα της συνάρτησης πυκνότητας της κατανομής βήτα για (γ) r=0.5, s=0.5, 2.0, 10.0.

Επομένως, είναι δυνατόν σε πολλές περιπτώσεις να προσαρμοσθεί η κατάλληλη θεωρητική πυκνότητα στα εμπειρικά δεδομένα.

Πολλές από τις μαθηματικές ιδιότητες της κατανομής Βήτα στηρίζονται στο ορισμένο ολοκλήρωμα

$$B(r,s) = \frac{\Gamma(r)\ \Gamma(s)}{\Gamma(r+s)} = \int_0^1 t^{r-1}(1-t)^{s-1}dt \qquad t\ >0,\ s>0$$

Ορισμός: Θα λέμε ότι η (απόλυτα) συνεχής τυχαία μεταβλητή X, με εύρος τιμών (0,1), ακολουθεί την *κατανομή Βήτα με θετικές παραμέτρους* (r, s) και θα συμβολίζουμε με $X \sim \text{beta}\ (r,s)$, αν

$$f_X(t) = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}\, t^{r-1}\,(1-t)^{s-1} \qquad 0<t<1,\ r>0,\ \ s>0$$

Αν $r \geq 1,\ s \geq 1$, τα άκρα του διαστήματος περιλαμβάνονται στο εύρος των τιμών της κατανομής Βήτα. Αν $r \geq 2,\ s \geq 2$, η συνάρτηση πυκνότητας πιθανότητας παρουσιάζει μέγιστο στο σημείο

$$t = t_0 = \frac{r-1}{(r-1)+(s-1)} = \frac{r-1}{r+s-2}$$

Αν οι παράμετροι r, s έχουν θετικές ακέραιες τιμές, η συνάρτηση πυκνότητας πιθανότητας μπορεί να εκφρασθεί μέσω ενός πολυωνύμου στο διάστημα [0,1], και έχει την τιμή 0 οπουδήποτε αλλού. Για παράδειγμα,

r	s	$f_X(t),\ 0 \le t \le 1$	Μέγιστο στο σημείο
1	1	1	
1	2	2(1-t)	
1	5	$5(1-t)^4$	
2	1	2t	
2	2	6t(1-t)	1/2
2	5	$30t(1-t)^4$	1/5

Η συνάρτηση κατανομής F_X δίνεται από τον τύπο

$$F_X(t) = \frac{1}{B(r,\ s)} \int_0^t u^{r-1}(1-u)^{s-1}\,du\ , \qquad 0 < t < 1$$

Για την περίπτωση στην οποία οι παράμετροι r, s έχουν θετικές ακέραιες τιμές, η $F_X(t)$ εκφράζεται μέσω ενός πολυωνύμου ως προς t για 0 < t < 1. Έτσι, για r = 2, s = 2,

$$F_X(t) = \int_0^t (6u - 6u^2)\,du = 3t^2 - 3t^3, \qquad 0 \le t \le 1$$

Στην γενική περίπτωση, το παραπάνω ολοκλήρωμα ορίζει την *ελλιπή συνάρτηση βήτα (incomplete beta function)*. Η συνάρτηση αυτή έχει ευρέως μελετηθεί και μία πληθώρα ειδικών περιπτώσεων, προσεγγίσεων και σχέσεων με άλλες κατανομές έχει βρεθεί.

Ιδιότητες: Μπορεί να αποδειχθεί ότι όταν $r \ge 0,\ s \ge 0$,

$$B(r,\ s) = \int_0^1 u^{r-1}(1-u)^{s-1}\,du$$

Επομένως, όταν η τυχαία μεταβλητή X ακολουθεί την κατανομή Βήτα με παραμέτρους r και s,

$$E(X^k) = \frac{1}{B(r,\ s)}\int_0^1 x^{r+k-1}(1-x)^{s-1}dx = \frac{B(r+k,\ s)}{B(r,\ s)} = \frac{\Gamma(r+k)\Gamma(r+s)}{\Gamma(r)\Gamma(r+s+k)}, \quad k \geq 0$$

Εφαρμόζοντας το αποτέλεσμα αυτό όταν k=0 οδηγεί στο συμπέρασμα ότι

$$\int_0^1 f_x(x)dx = 1$$

Το παραπάνω αποτέλεσμα, επειδή $f_X(x)\geq 0$, για όλες τις τιμές x, συνεπάγεται ότι η συνάρτηση $f_X(x)$ είναι μία καλώς ορισμένη συνάρτηση πυκνότητας πιθανότητας.

Θέτοντας k = 1 και k = 2 στον τύπο που προσδιορίζει την μορφή της $E(X^k)$, έχουμε

$$E(X) = \frac{\Gamma(r+1)}{\Gamma(r)}\frac{\Gamma(r+s)}{\Gamma(r+s+1)} = \frac{r}{r+s}$$

και

$$E(X^2) = \frac{\Gamma(r+2)}{\Gamma(r)}\frac{\Gamma(r+s)}{\Gamma(r+s+2)} = \frac{(r+1)(r)}{(r+s+1)(r+s)}$$

Επομένως,

$$\mu_X = E(X) = \frac{r}{r+s}, \qquad \sigma_X^2 = E(X^2) - (\mu_X)^2 = \frac{rs}{(r+s)^2(r+s+1)}$$

Είναι ενδιαφέρον να παρατηρηθεί ότι αν η μεταβλητή Y έχει την κατανομή Βήτα με παραμέτρους r και s, τότε η μεταβλητή 1-Y έχει επίσης την κατανομή Βήτα με παραμέτρους s και r.

Πράγματι, αν U = 1–Y, τότε επειδή B(r, s)=B(s, r), η συνάρτηση πυκνότητας πιθανότητας της U είναι η

$$f_U(u) = \frac{1}{|-1|}f_Y(1-u) = \begin{cases} \dfrac{(1-u)^{r-1}u^{s-1}}{B(r,s)}, & \text{αν } 0 \leq 1-u \leq 1, \\ 0, & \text{διαφορετικά} \end{cases}$$

$$= \begin{cases} \dfrac{u^{s-1}(1-u)^{r-1}}{B(s,r)}, & \text{αν } 0 \leq u \leq 1, \\ 0, & \text{διαφορετικά} \end{cases}$$

η οποία είναι ακριβώς η συνάρτηση πυκνότητας πιθανότητας μίας κατανομής Βήτα με παραμέτρους s και r.

Σχέση μεταξύ της κατανομής Βήτα και της Διωνυμικής κατανομής

Μπορεί να αποδειχθεί ότι αν η τυχαία μεταβλητή Y έχει την κατανομή Βήτα με παραμέτρους r και s, όπου r και s είναι θετικοί ακέραιοι και η τυχαία μεταβλητή X ακολουθεί την διωνυμική κατανομή με παραμέτρους n = r + s – 1 και p = y, τότε για $0 \leq y \leq 1$,

$$F_Y(y) = P\{r \leq X \leq r+s-1\} = 1 - F_X(r-1)$$

Η ιδιότητα αυτή έχει πολλές εφαρμογές.

Παράδειγμα: Έστω Y το ποσοστό κορεσμού διαλυμένου οξυγόνου σε ένα συγκεκριμένο σημείο ενός ποταμού. Η κατανομή της μεταβλητής Y σε καθαρό νερό ακολουθεί την κατανομή Βήτα με παραμέτρους r = 3 και s = 2. Ένα σύστημα συναγερμού για την ποιότητα του νερού ειδοποιεί όταν η παρατηρούμενη τιμή της μεταβλητής Y πέσει κάτω από ένα συγκεκριμένο επίπεδο y^*. Το σύστημα αυτό δεν πρέπει να δίνει λανθασμένη ένδειξη παρουσίας μη καθαρού νερού πολύ συχνά όταν το νερό είναι στην πραγματικότητα καθαρό. Αν $y^* = 0.40$, τότε από την παραπάνω σχέση που συνδέει τις συναρτήσεις κατανομών των μεταβλητών Y και X και τον πίνακα της διωνυμικής κατανομής, για n = 3 + 2 – 1 = 4 και p = 0.40, έχουμε

$$P(\text{λανθασμένο σύνθημα κινδύνου}) = P\{Y \leq 0.40\} = F_Y(0.40)$$
$$= P\{3 \leq X \leq 3+2-1\}$$
$$= P\{X=3\} + P\{X=4\} = 0.15360 + 0.02560 = 0.17920$$

Αν αυτή η πιθανότητα λανθασμένου συνθήματος κινδύνου είναι πολύ μεγάλη, η τιμή y^* μπορεί να ελαττωθεί.

Για $y^* = 0.20$

$$P\{Y \le 0.20\} = F_Y(0.20) = P\{3 \le X \le 4\} = 0.02560 + 0.00160$$
$$= 0.02720$$

η οποία είναι μία ικανοποιητικά χαμηλή τιμή.