

Συνεχείς κατανομές πιθανοτήτων εισαγωγή

--- κανονική κατανομή

συνταίριασμα μεταξύ δείγματος και
θεωρητικών κατανομών

Δρ Μ.Σπηλιώτη

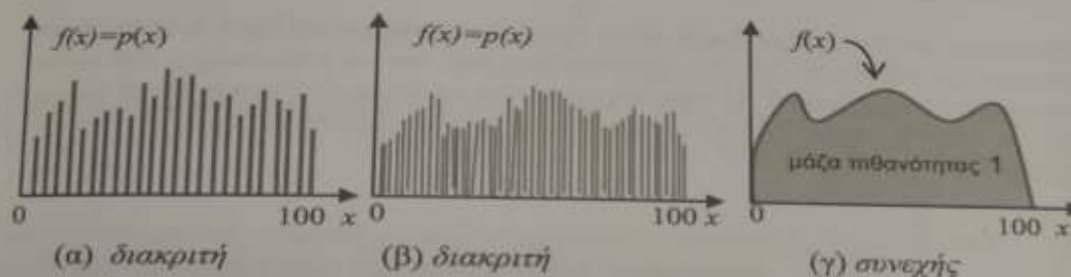
Αναπληρωτής Καθηγητής ΔΠΘ



Συνεχής τ.μ., συνάρτηση πυκνότητας
πιθανότητας, η κανονική κατανομή

Ζιούτας Γ, 2012. Πιθανότητες και Στατιστική για Μηχανικούς. Εκδόσεις Σοφία

Εστω ότι έχουμε μία διακριτή τυχαία μεταβλητή X η οποία μπορεί να πάρει 25 διακριτές τιμές από το 0 μέχρι το 100 με συνάρτηση μάζας πιθανότητας όπως φαίνεται στο Σχήμα 4.5 (α). Ας πούμε τώρα ότι η τυχαία μεταβλητή X μπορεί να πάρει περισσότερες τιμές στο διάστημα $0 \leq x \leq 100$, όπου ο αριθμός τιμών είναι μεγάλος αλλά πεπερασμένος, όπως 1, 2, 3, ..., 98, 99, 100. Η μάζα πιθανότητας των 25 μπάρων απλώνεται τώρα σε περισσότερα σημεία της X , όπως παριστάνεται στην περίπτωση (β) του Σχήματος 4.5.



Σχήμα 4.5 (α), (β), συναρτήσεις μάζας πιθανότητας, (γ) συνάρτηση πυκνότητας πιθανότητας

Σε κάθε μία από τις δυνατές τιμές x_i αντιστοιχεί μία πιθανότητα $p_i = p(x_i) = P(X=x_i)$, όπου $i=1, 2, \dots, 100$, των οποίων το άθροισμα ισούται με την μονάδα. Αναφέραμε παραπάνω ότι η τυχαία μεταβλητή X ενδέχεται να πάρει οιαδήποτε τιμή μέσα στο διάστημα, $0 \leq x \leq 100$. Αφού οι τιμές που μπορεί να πάρει η X είναι άπειρες μη μετρήσιμες, τι θα συμβεί με τις πιθανότητες $p_i = p(x_i)$; Οι πιθανότητες p_i είναι άπειρες τον αριθμό και το άθροισμά τους θα οδηγούσε σε τιμή μεγαλύτερη της μονάδας. Οι πιθανότητες αυτές έχουν μηδενική τιμή και δεν έχουν νόημα. Εκείνο που πρέπει να κάνουμε είναι να σκορπίσουμε την μάζα των πιθανοτήτων $p(x_i)$ που ορίζονται μόνο για τις τιμές $x_1, x_2, x_3, \dots, x_n$ σε όλο το διάστημα, $0 \leq x \leq 100$, όπως φαίνεται στο Σχήμα 4.5 (γ). Προκύπτει έτσι μία συνάρτηση f ορισμένη (για το παρών παράδειγμα) για όλες τις τιμές x , $0 \leq x \leq 100$, η οποία εκφράζει την πυκνότητα της μάζας πιθανότητας σε όλο το πεδίο τιμών της X .

$$P(a < X \leq b) = \int_a^b f_X(x) dx \quad (3.3)$$

Με βάση την (3.3), η συνάρτηση κατανομής είναι

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(\xi) d\xi \quad (3.4)$$

και αν η πρώτη παράγωγος της $F_X(x)$ υπάρχει, τότε

$$\left(f_X(x) = \frac{dF_X(x)}{dx} \right) \quad (3.5)$$

Πρέπει ίσως να επαναληφθεί ότι η συνάρτηση $f_X(x)$ δεν εκφράζει πιθανότητες. Όμως το $f_X(x) \cdot dx = P(x < X \leq x + dx)$ είναι η πιθανότητα ότι οι τιμές της X θα είναι στο διάστημα $[x, x + dx]$.

Συνάρτηση πυκνότητα πιθανότητας (για συνεχής τ.μ.)

Ορισμός: Ονομάζουμε συνάρτηση πυκνότητας
πιθανότητας, μιας συνεχούς τυχαίας μεταβλητής X τη
συνάρτηση $f_X(x)$, για την οποία ισχύουν:

$$(1) f_X(x) \geq 0, \quad \forall x \in \mathbb{R}. \quad (\text{θετική τιμή})$$

$$(2) \int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

(3) η πιθανότητα του γεγονότος $X \in (a, b)$:

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

Η συνάρτηση $f(x)$ δεν αποδίδει πιθανότητα.

Οπώς: $f_X(x)dx = P(x \leq X \leq x+dx)$ δηλ.

$f_X(x)dx =$ η πιθανότητα να βρίσκεται η X σε έναν από τους διαστήματα $[x, x+dx)$

ΘΕΩΡΗΤΙΚΗ ΠΙΘΑΝΟΤΗΤΑ

(Συνάρτηση Κατανομής και αθροιστική πιθανότητα)

Συνάρτηση πυκνότητας πιθανότητας

X τυχαία μεταβλητή

Συνάρτηση πυκνότητας πιθανότητας
(συνάρτηση κατανομής)

$$f_X(X) = \frac{dF_X(x)}{dx}$$

Πιθανότητα μη υπέρβασης

Η πιθανότητα η τυχαία
μεταβλητή να είναι μικρότερη ή
ίση της δεδομένης τιμής x

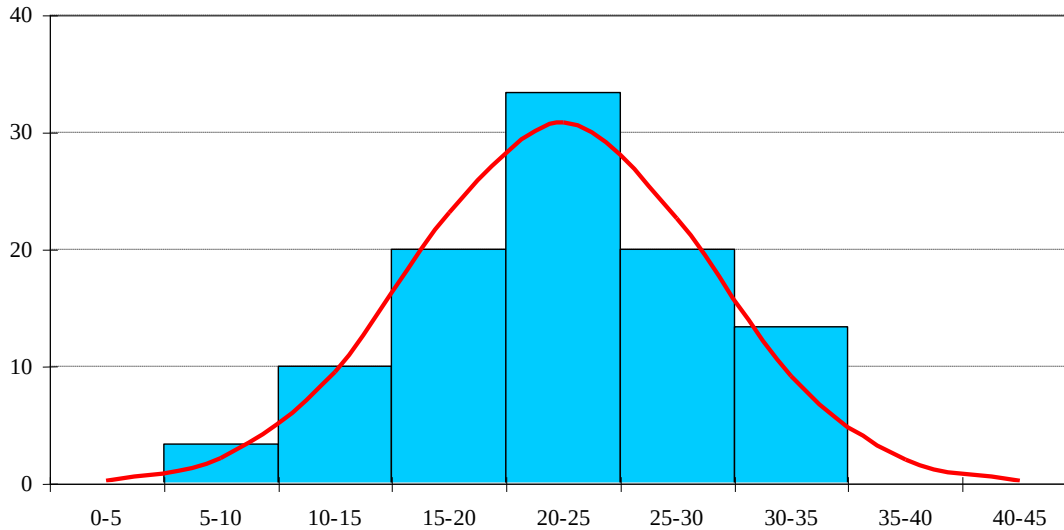
$$F_X(x) = P(X \leq x) \text{ (αθροιστική Πιθανότητα)}$$
$$0 = F_X(-\infty) \leq F_X(x) \leq F_X(+\infty) = 1$$

Πιθανότητα υπέρβασης

Η πιθανότητα η τυχαία μεταβλητή να
είναι μεγαλύτερη της δεδομένης τιμής x

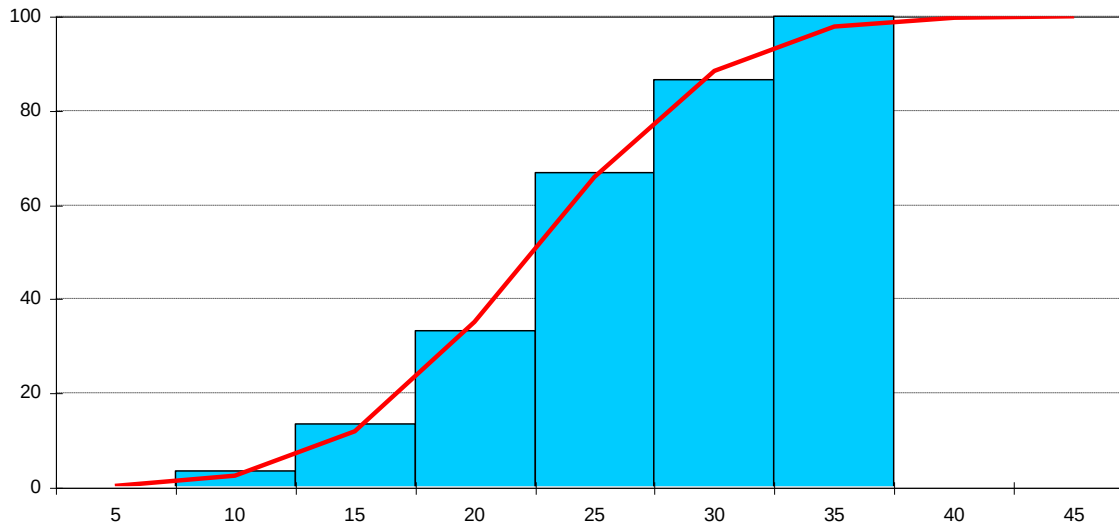
$$F_{1X} = P(X \geq x) = 1 - F_X(x)$$

Διάγραμμα συνάρτησης πυκνότητας
πιθανότητας (συνάρτηση κατανομής)



Η συνάρτηση πυκνότητας
πιθανότητας ισούται με την
πρώτη παράγωγο της
αθροιστικής πιθανότητας

Διάγραμμα αθροιστικής πιθανότητας



Η αθροιστική πιθανότητα
δηλώνει την πιθανότητα όπου
μια τυχαία μεταβλητή είναι
ίση ή μικρότερη από κάποια
δοθείσα τιμή

Βασικές έννοιες (συνεχείς και διακριτές κατανομές)

Οι τιμές των Υδρολογικών δεδομένων είναι δυνατόν να θεωρηθούν τυχαίες οπότε μελετώνται σαν Στατιστικές μεταβλητές.

Πληθυσμός στη Στατιστική είναι ένα σύνολο στοιχείων με κοινές ιδιότητες π.χ. τα ετήσια ύψη βροχής ενός βροχομετρικού σταθμού.

Δείγμα είναι ένα τμήμα του πληθυσμού.

Οι μεταβλητές διακρίνονται σε

διακεκριμένες όπως π.χ. ο αριθμός των εμφανίσεων μιας τιμής της παροχής ενός ποταμού σε μία θέση μεγαλύτερης από μια συγκεκριμένη τιμή κατά τη διάρκεια μιας χρονικής περιόδου και

συνεχείς όπως π.χ. οι τιμές της μέσης ημερήσιας παροχής κατά τη διάρκεια μιας χρονικής περιόδου.

➤ Σε πολλά προβλήματα (υδρολογίας, υδρογεωλογίας, μετεωρολογίας κ.α) **οι μεταβλητές θεωρούνται τυχαίες** (π.χ. απορροή, βροχόπτωση κτλ.).

➤ Κάθε τυχαία μεταβλητή μπορεί να περιγραφεί από μια συνάρτηση πυκνότητας πιθανότητας και την αντίστοιχη αθροιστική πιθανότητα κάποια κατανομής.

➤ Η εύρεση της κατανομής που ακολουθεί η τυχαία μεταβλητή είναι κρίσιμης σημασίας διότι γίνεται εφικτή η πρόβλεψη εμφάνισης ή μη μιας τιμής της μεταβλητής κάτω από μια πιθανότητα.

➤ Η εύρεση της κατανομής γίνεται με το ταίριασμα της θεωρητικής πιθανότητας (που υποθέτουμε ότι ακολουθεί το δείγμα) με την εμπειρική (που υπολογίζεται βάσει Weibull).

➤ Για το ταίριασμα αυτό αξιοποιούνται εργαλεία της **Στατιστικής**. Υπολογίζονται οι στατιστικές παράμετροι του μέσου όρου, της τυπικής απόκλισης, του συντελεστή μεταβλητότητας και του συντελεστή ασυμμετρίας από το δείγμα.

Κανονική κατανομή (Normal distribution)

Συνάρτηση πυκνότητας πιθανότητας:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.32)$$

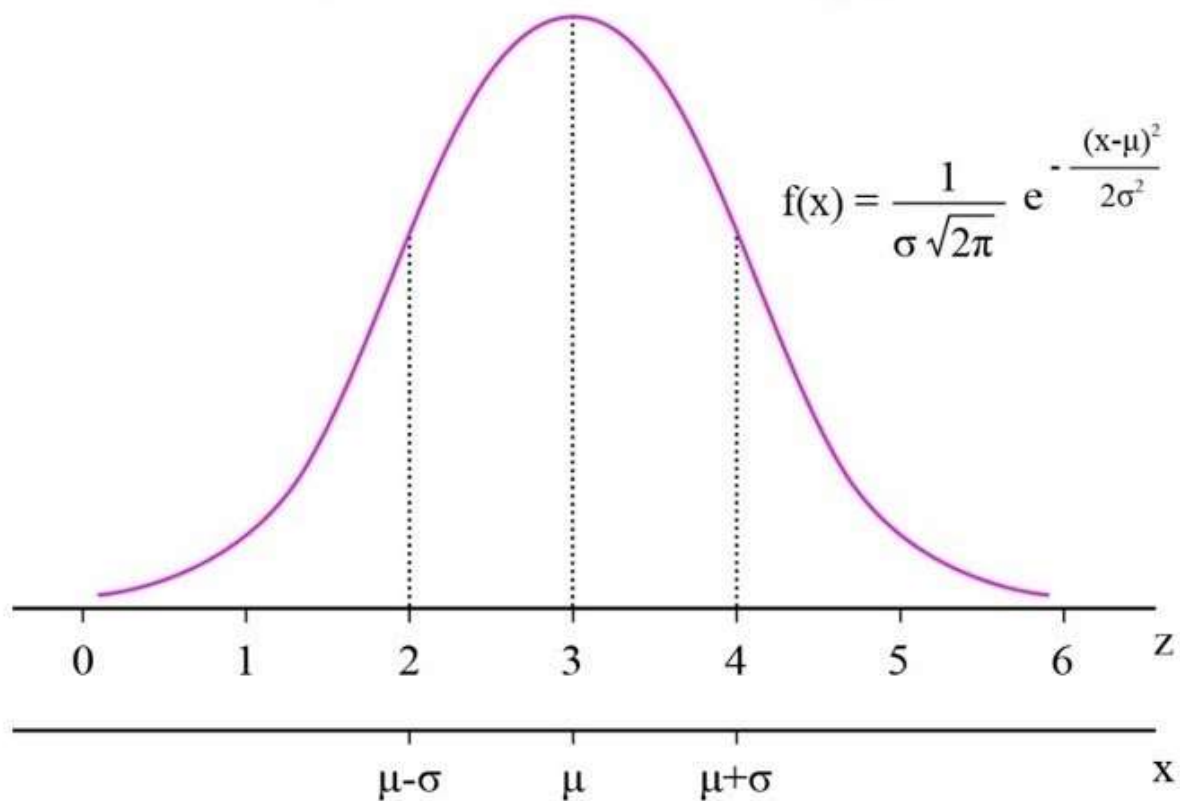
Αθροιστική συνάρτηση πιθανότητας:

$$p(X \leq x) = F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x p(x) dx \quad (2.33)$$

$$\text{αν } z = \frac{x-\mu}{\sigma} \quad p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} = \sigma p(x) \quad (2.34)$$

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z p(z) dz \quad (2.35)$$

Κανονική κατανομή



Κανονική κατανομή συνάρτηση πυκνότητας πιθανότητας

$f(X)$

μ , μετακίνηση καμπύλης
αριστερά η δεξιά

σ αυξάνει η μικραίνει το
πλάτος



Συνάρτηση πυκνότητας πιθανότητας κανονική κατανομή(pdf)

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

Note constants:

$\pi=3.14159$

$e=2.71828$

Παράμετροι, από
στατιστικό δείγμα

Συνάρτηση αθροιστικής
πιθανότητας κανονική
κατανομή(cpdf)

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

πιθ, τιμές μικρότερες ή ίσες του x

Μεταβλητή (Z)

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$



$$Z = \frac{X - \mu}{\sigma}$$

$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-1/2 z^2} dz = \Phi(z)$$

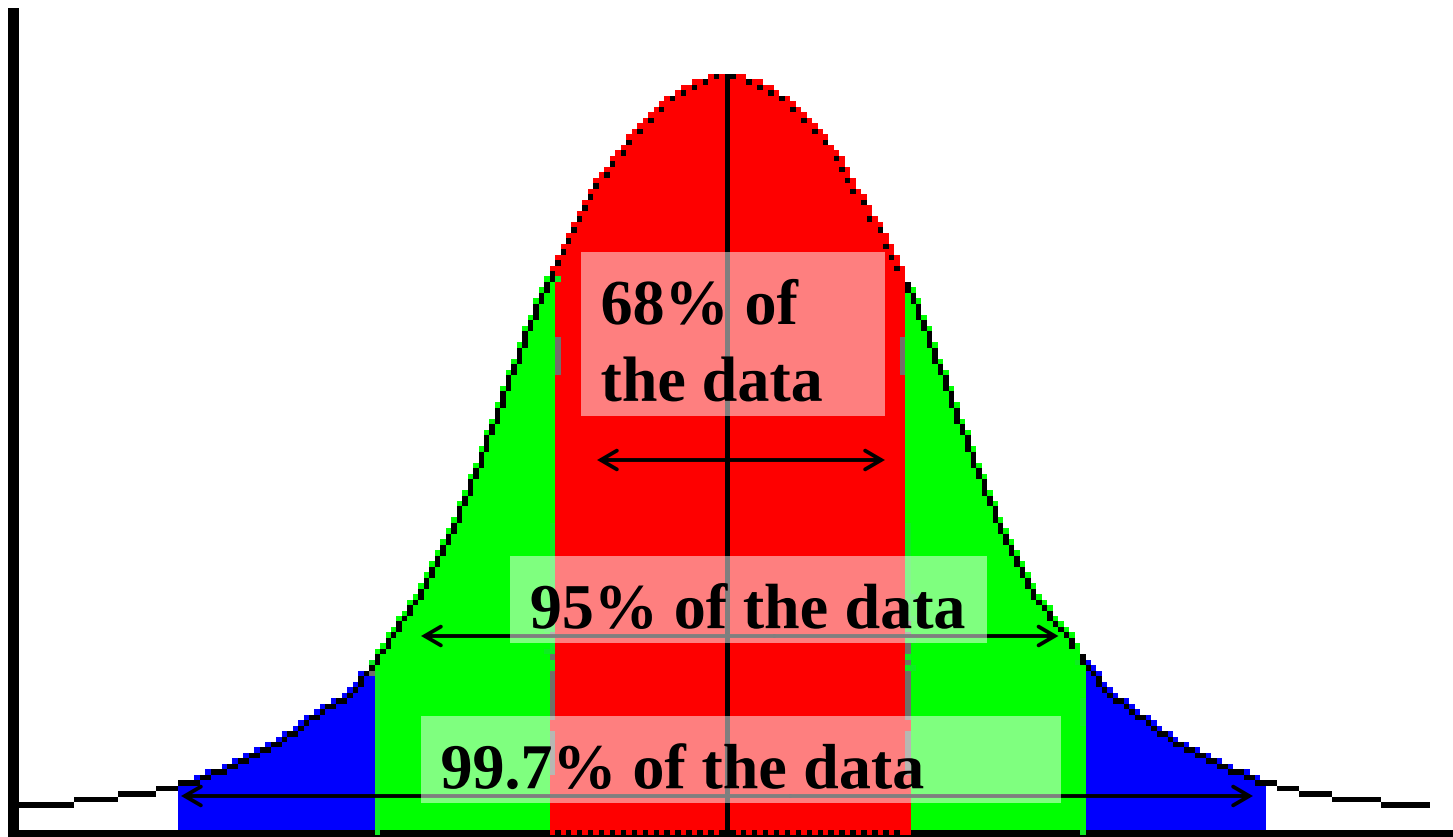
πίνακες

Βασική ιδιότητα

Όλα τα ενδεχόμενα , πιθανότητα 1!

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

68-95-99.7 Κανόνας



68-95-99.7 κανόνας...

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

$$\Phi\left(\frac{\mu+\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-\sigma-\mu}{\sigma}\right) = \Phi(1) - \Phi(-1) = .682$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .95$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .997$$

Συνάρτηση αθροιστικής
πιθανότητας κανονική
κατανομή(cpdf)

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx =$$

πιθ, τιμές μικρότερες ή ίσες του x

Μεταβλητή (Z)

$$P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

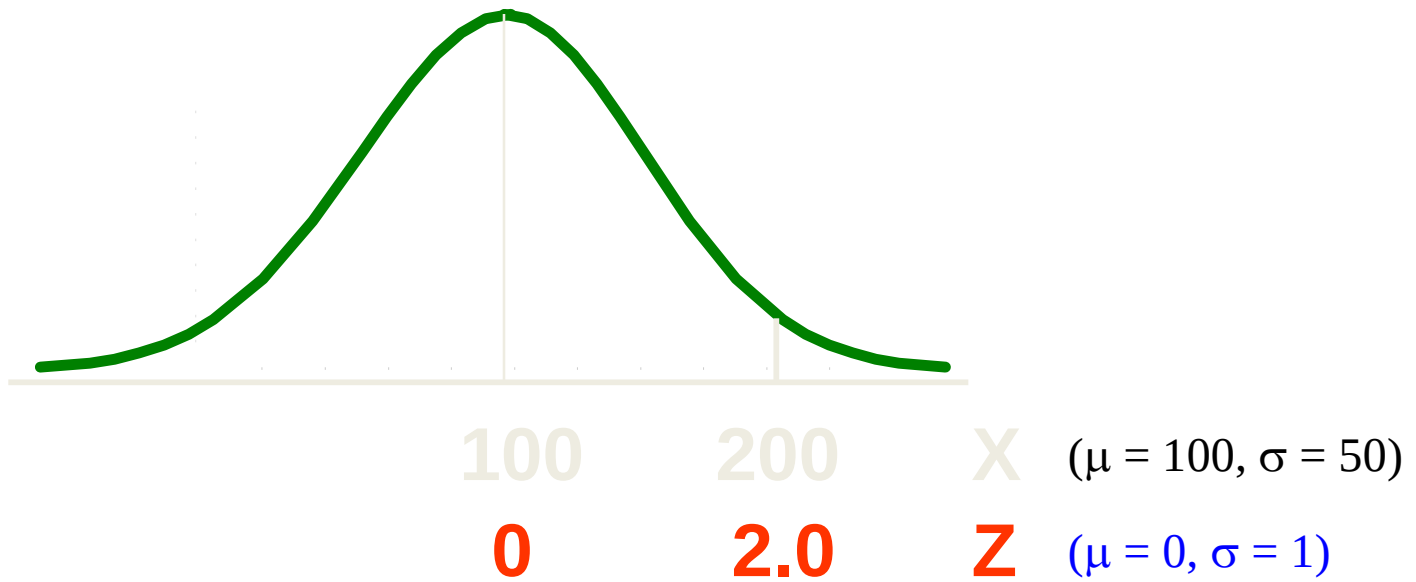


$$Z = \frac{x - \mu}{\sigma}$$

$$P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-1/2 z^2} dz$$

πίνακες

Σύγκριση X και Z



προβληματάκι

b. Πιθανότητα $X \leq 120$?

$$Z = \frac{120 - 109}{13} = .85$$

$$P(X \leq 120) = P(Z \leq .85) = .8023 = 80.23\%$$

ΠΑΡΑΔΕΙΓΜΑ 3.6

Ἡ ὁλική ἐτήσια βροχόπτωση σὲ μιὰ περιοχή εἶναι κανονικὴ τ.μ.: $N(60\text{cm}, 15\text{cm})$.

(α) Ποιὰ εἶναι ἡ πιθανότητα ὅτι τὸν ἐπόμενο χρόνο ἡ βροχόπτωση θὰ εἶναι μεταξύ 40 καὶ 70cm;

Σύμφωνα μὲ τὶς Σχέσεις (3.28) καὶ 3.27β, ἡ πιθανότητα εἶναι

$$\begin{aligned} P(40 < X \leq 70) &= \Phi\left(\frac{70-60}{15}\right) - \Phi\left(\frac{40-60}{15}\right) \\ &= \Phi(0.67) - \Phi(-1.33) \end{aligned}$$

(β) Ποιὰ εἶναι ἡ πιθανότητα ὅτι ἡ ἐτήσια βροχόπτωση θὰ εἶναι τοὐλάχιστον 30cm;

$$P(X \geq 30) = 1 - P(X \leq 30) = 1 - 0.02275 = 0.97725$$

(γ) Ποιὰ εἶναι ἡ τιμὴ τοῦ 10^{ου} ἐκατοστημίου τῆς ἐτήσιας βροχόπτωσης (δηλαδή, ἡ τιμὴ στὴν ὁποία ἡ ἀθροιστικὴ πιθανότητα εἶναι 10%);

Συμβολίζοντας τὴ τιμὴ τοῦ 10^{ου} ἐκατοστημίου μὲ $x_{.10}$, ἔχουμε

$$P(X \leq x_{.10}) = 0.10,$$

ὁρα

$$\Phi\left(\frac{x_{.10} - 60}{15}\right) = 0.10$$

Μὲ τὴ βοήθεια τῆς (3.27γ),

$$\frac{x_{.10} - 60}{15} = \Phi^{-1}(0.10) = -\Phi^{-1}(0.90) = -1.28$$

ὁπότε

$$x_{.10} = 60 - 1.28(15) = 40.8\text{cm}$$

Alferdo, H., Ang, S., Tang, H., W, 1984. Εφαρμογές πιθανοτήτων και στατιστικής

στη μελέτη και προγραμματισμό τεχνικών έργων. Μετάφραση – επιμέλεια: Παναγιωτακόπουλος Δ., Καθηγητὴς Πολυτεχνικῆς Σχολῆς ΔΠΘ, Δημοκρίτειο Πανεπιστήμιο Θράκης, Ξάνθη 1984, pp. 94-95.

NORM.DIST (Συνάρτηση NORM.DIST)

Επιστρέφει την κανονική κατανομή για καθορισμένη μέση τιμή και τυπική απόκλιση. Η συνάρτηση αυτή έχει πολύ μεγάλο εύρος εφαρμογής στη στατιστική, καθώς και στον έλεγχο υποθέσεων.

Σύνταξη

NORM.DIST(x;μέση_τιμή;τυπική_απόκλιση;αθροιστική)

Η σύνταξη της συνάρτησης NORM.DIST περιλαμβάνει τα παρακάτω ορίσματα:

- **X** Υποχρεωτικό. Η τιμή για την οποία θέλετε την κατανομή.
- **Μέση_τιμή** Υποχρεωτικό. Ο αριθμητικός μέσος της κατανομής.
- **Τυπική_απόκλιση** Υποχρεωτικό. Η τυπική απόκλιση της κατανομής.
- **Αθροιστική** Υποχρεωτικό. Μια λογική τιμή που καθορίζει το είδος της συνάρτησης. Εάν το όρισμα αθροιστική είναι TRUE, η συνάρτηση NORM. Η συνάρτηση DIST επιστρέφει τη συνάρτηση αθροιστικής κατανομής. Εάν είναι FALSE, επιστρέφει τη συνάρτηση πυκνότητας πιθανότητας.

Παρατηρήσεις

- Εάν το όρισμα μέση_τιμή ή τυπική_απόκλιση δεν είναι αριθμητικό, η συνάρτηση NORM.DIST επιστρέφει #ΤΙΜΗ! ως τιμή σφάλματος.
- Εάν τυπική_απόκλιση ≤ 0 , η συνάρτηση NORM.DIST επιστρέφει #ΑΠΙΘ! ως τιμή σφάλματος.
- Εάν μέση_τιμή = 0, τυπική_απόκλιση = 1 και αθροιστική = TRUE, η συνάρτηση NORM.DIST επιστρέφει την τυπική κανονική κατανομή, τη συνάρτηση NORM.S.DIST.

mo	60		0.022 75		0.977 25
st	15				
x	30				

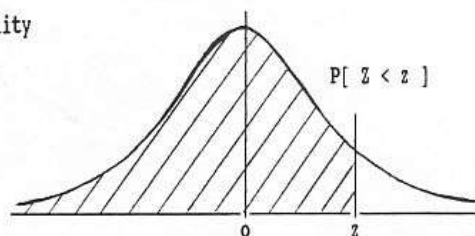
mo	60
st	15
x	40.77673

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability
up to the standardised normal value z
i.e.

$$P[Z < z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Αντίστροφο πρόβλημα

- Μέθοδος του παράγοντα συχνότητας

Σύμφωνα με τον Chow 1951 για πολλές θεωρητικές κατανομές πιθανότητας μπορεί να προσδιοριστεί η τιμή της τυχαίας μεταβλητής με βάση την αθροιστική κατανομή πιθανότητας σύμφωνα με την παρακάτω εξίσωση:

$$x_T = \mu + \sigma K_T \quad (1)$$

όπου μ ο μέσος όρος, σ η τυπική απόκλιση και K_T ο παράγοντας συχνότητας, που καθορίζεται από την περίοδο επαναφοράς και προκύπτει από αλγεβρικούς μετασχηματισμούς για κάθε κατανομή πυκνότητας πιθανότητας.

Κανονική κατανομή, $K_T = z$ $Z = \frac{x - \mu}{\sigma} \Leftrightarrow x = \mu + \sigma Z$

Περίοδος επαναφοράς

Μέγιστα

Για κατανομές μεγίστων, για κάθε πιθανότητα υπέρβασης μπορεί να ορισθεί η συνακόλουθη περίοδος επαναφοράς T , που είναι το μέσο χρονικό διάστημα, που η τυχαία (υδρολογική) μεταβλητή θα εμφανισθεί μόνο μία φορά με τιμή μεγαλύτερη ή ίση της δοθείσας (Τσακίρης 2012):

$$P(Q \geq q) = 1 - P(Q \leq q) = \frac{1}{T}$$

Ελάχιστα

Για κατανομές ελαχίστων, για κάθε πιθανότητα υπέρβασης μπορεί να ορισθεί η συνακόλουθη περίοδος επαναφοράς T , που είναι το μέσο χρονικό διάστημα, που η τυχαία (υδρολογική) μεταβλητή θα εμφανισθεί μόνο μία φορά με τιμή μικρότερη ή ίση της δοθείσας (Τσακίρης 2012):

$$P(Q \leq q) = \frac{1}{T}$$

Πίνακας με τις τιμές της μεταβλητής Z . Στην κανονική κατανομή ο παράγοντας συχνότητας K συμπίπτει με τη μεταβλητή Z .

Ο παράγοντας συχνότητας για την κανονική κατανομή είναι ίσος με την ανηγμένη μεταβλητή, $k_T = z$.

Πιν. 2.1: Τιμές του z για δοσμένες τιμές της αθροιστικής πιθανότητας $P(z)$ της Τυπικής Κανονικής Κατανομής

$P(z)$ %	z	$P(z)$ %	z	$P(z)$ %	z	$P(z)$ %	z	$P(z)$ %	z	$P(z)$ %	z
1	-2.326	21	-0.806	41	-0.228	61	0.279	81	0.878	99.1	2.366
2	-2.054	22	-0.772	42	-0.202	62	0.305	82	0.915	99.2	2.409
3	-1.881	23	-0.739	43	-0.176	63	0.332	83	0.954	99.3	2.457
4	-1.751	24	-0.706	44	-0.151	64	0.358	84	0.994	99.4	2.512
5	-1.645	25	-0.674	45	-0.126	65	0.385	85	1.036	99.5	2.576
6	-1.555	26	-0.643	46	-0.100	66	0.412	86	1.080	99.6	2.652
7	-1.476	27	-0.613	47	-0.075	67	0.440	87	1.126	99.7	2.748
8	-1.405	28	-0.583	48	-0.050	68	0.467	88	1.175	99.8	2.878
9	-1.341	29	-0.553	49	-0.025	69	0.496	89	1.227	99.9	3.090
10	-1.382	30	-0.524	50	0.000	70	0.524	90	1.282		
11	-1.227	31	-0.496	51	0.025	71	0.553	91	1.341	99.91	3.121
12	-1.175	32	-0.468	52	0.050	72	0.583	92	1.405	99.92	3.156
13	-1.126	33	-0.440	53	0.075	73	0.613	93	1.476	99.93	3.195
14	-1.080	34	-0.412	54	0.100	74	0.643	94	1.555	99.94	3.239
15	-1.036	35	-0.385	55	0.126	75	0.674	95	1.645	99.95	3.291
16	-0.994	36	-0.358	56	0.151	76	0.706	96	1.751	99.96	3.353
17	-0.954	37	-0.332	57	0.176	77	0.739	97	1.881	99.97	3.432
18	-0.915	38	-0.305	58	0.202	78	0.772	98	2.054	99.98	3.540
19	-0.878	39	-0.279	59	0.228	79	0.806	99	2.326	99.99	3.719
20	-0.842	40	-0.253	60	0.253	80	0.842				

Παράδειγμα:

Αν σε ένα πληθυσμό $\mu = 20$ και $\sigma = 5$ να ευρεθεί η πιθανότητα $X < 2$.
Υπολογίζεται η ανοιγμένη μεταβλητή z :

$$z = (x - \mu) / \sigma = (2 - 20) / 5 = -3.6$$

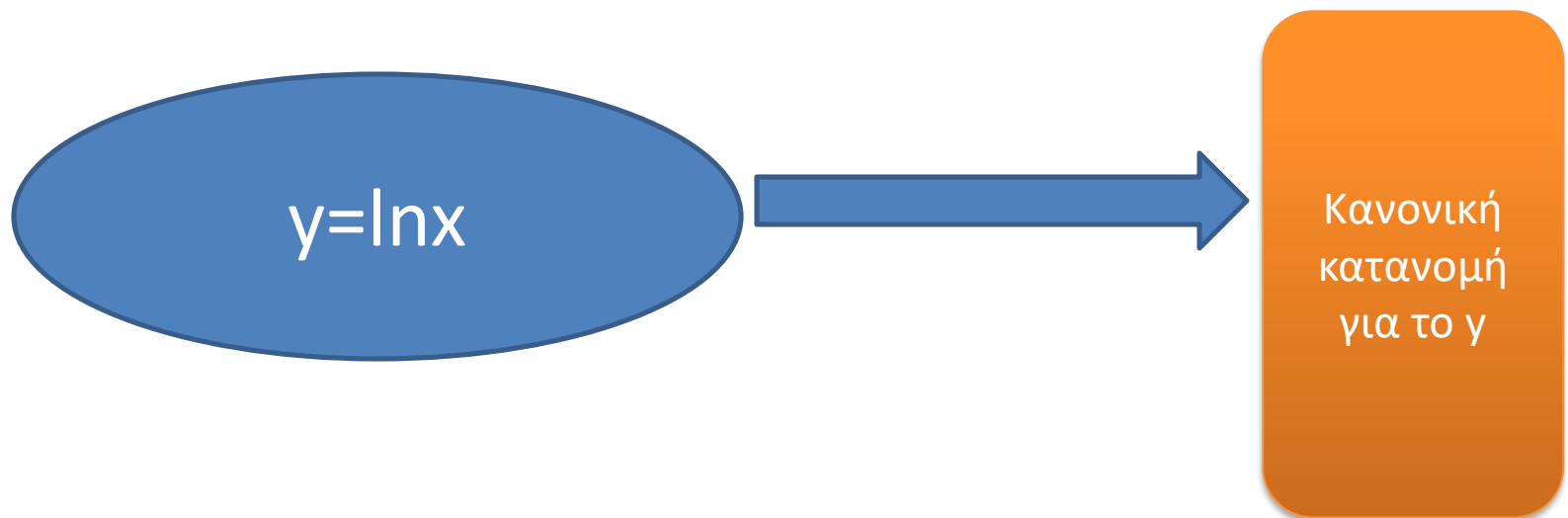
Από τους πίνακες της Κανονικής κατανομής (Παραρτ. Πιν. 1) για $z = -3.6$

Λογαριθμικές Κατανομές (περίπτωση Λογαριθμοκανονικής κατανομής)

Λογαριθμοκανονική κατανομή (Log-normal ή κατανομή Galton)

Στην κατανομή αυτή η κανονική κατανομή προσαρμόζεται στους Νεπέριους λογαρίθμους της στατιστικής μεταβλητής οπότε αν x_i είναι η στατιστική μεταβλητή τότε στην κανονική κατανομή προσαρμόζεται η μεταβλητή $y_i = \ln x_i$.

Οι αντίστοιχοι υπολογισμοί γίνονται όπως στην κατανομή Log Pearson III.



Οι στατιστικές παράμετροι της αρχικής μεταβλητής x συνδέονται με τις αντίστοιχες παραμέτρους της μεταβλητής y με τις ακόλουθες σχέσεις

$$\mu = e^{\mu_y + \frac{1}{2} \sigma_y^2} \quad (4.37)$$

$$\sigma = \mu \left[e^{\sigma_y^2} - 1 \right]^{1/2} \quad (4.38)$$

$$C_v = \left[e^{\sigma_y^2} - 1 \right]^{1/2} \quad (4.39)$$

$$C_s = 3 C_v + C_v^3 \quad (4.40)$$

Στατιστικό Δείγμα

Αμερόληπτοι εκτιμητές

Στατιστικό δείγμα - αφερόμενοι ευρεσιμότητες:

• Μέσος όρος

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N}$$

x_i : μέτρηση
 N : πλήθος δείγματος

• Τυπική απόκλιση

$$\hat{S} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1}}$$

• Άλλοι αφερόμενοι ευρεσιμότητες

Συντελεστής
ασυμμετρίας

$$C_3 \approx \frac{a}{\hat{S}^3} \rightarrow \text{ασυμμετρία}$$

$$= \frac{N}{(N-1)(N-2)} \sum (x_i - \bar{X})^3$$

- Ο μέσος όρος (average) στον πληθυσμό παρίσταται συνήθως με μ ενώ σε δείγμα με $\bar{x}$

$$\bar{x} = \frac{1}{N} \sum_i N_i x_i = \frac{1}{k} \sum x_i \quad (2.22)$$

- Η διάμεσος (median) είναι η τιμή που αντιστοιχεί στο 50% της κατανομής των πιθανοτήτων.
- Η πιθανότερη τιμή (mode) είναι η τιμή που εμφανίζεται πιο συχνά. Στις συνεχείς συναρτήσεις είναι η μέγιστη τιμή της συνάρτησης πυκνότητας πιθανότητας.
- Η μεταβλητότητα (variance) είναι:

$$\sigma^2 = \frac{1}{N} \sum (x_i - \bar{x})^2 \quad (2.23)$$

για μικρά δείγματα πολλαπλασιάζεται με την ποσότητα $\frac{N}{N-1}$

- Η τυπική απόκλιση (standard deviation) η οποία παρίσταται με σ ενώ σε δείγμα με s

$$\sigma = \sqrt{\mu_2} \quad s = \sqrt{\frac{1}{N-1} \sum N_i (x_i - \bar{x})^2} \quad (2.24)$$

- Τυπικό σφάλμα (standard error) ονομάζεται η ποσότητα: $\frac{s}{\sqrt{N}}$
- Ο συντελεστής μεταβλητότητας (coefficient of variation)

$$c_v = \frac{\sigma}{\mu} \approx \frac{s}{\bar{x}} \quad (2.25)$$

- Η ασυμμετρία $a = \mu_3$ για μικρά δείγματα είναι:

$$a = \frac{N}{(N-1)(N-2)} \sum (x_i - \bar{x})^3 \quad (2.26)$$

- Ο συντελεστής ασυμμετρίας

$$c_s = \frac{a}{\sigma^3} \approx \frac{a}{s^3} \quad (2.27)$$

- Κύρτωση:

Δύο κατανομές μπορούν να έχουν την ίδια μέση τιμή και τυπική απόκλιση αλλά να διαφέρουν μεταξύ τους ως προς την "κύρτωση" (οιχμηρότητα της καμπύλης συνωστίζεται πιθανότητας)

Υπάρχουν δύο μέτρα κύρτωσης.

Συντελεστή Κύρτωσης: K

$$K = \frac{1}{N-4} \sum_{i=1}^n (x_i - \bar{x})^4$$

για μικρά δείγματα γίνεται πολύ λανθασμένη
επί $\frac{N^3}{(N-1)(N-2)(N-3)}$

Για την κανονική κατανομή (ιδεώς κύρτος),

$K = 3$. (Μιρίκου, 1994)

• Η εντολή KURT του Excel υπολογίζει το ¹ (Ανάλυση και Απομείωση)
παρακάτω μέτρο: $\left\{ \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \right\}$

Συνταίριασμα μεταξύ στατιστικού δείγματος και θεωρητικών κατανομών

Γιατί χρησιμοποιώ θεωρητικές κατανομές πιθανοτήτων?

- Γιατί χρησιμοποιώ θεωρητικές κατανομές πιθανοτήτων και όχι την εμπειρική κατανομή?

Συνήθως, η περίοδος παραφοράς (ή η στο
αριθμός) είναι μεγαλύτερη (ή και πολύ μεγάλη)
από το μέγεθος του δείγματος. Επομένως επιφεύγουμε
σε θεωρητική συνάρτηση πυκνότητας πιθανότητας και
τις προσαρμόζουμε στο στατιστικό δείγμα

Πα Την 'Προσφεση' (ρύθμιση) των θεωρητικών
κατανομών πιθανότητας υπάρχουν διάφοροι μέθοδοι

1. η μέθοδος των ρατών
2. η μέθοδος της μέγιστης πιθανοφάνειας
3. ελάχιστος τετραγώνος
εάν

Ειδικά στην υβριδική χρο/και και οι

4. γραμμική μέθοδος
5. παράγοντα συχνότητας

Εκτίμηση παραμέτρων

Μέθοδος των ροπών

Μέσος Των ποσών

(A) Θεωρητικές Μεταβολές πιθανότητας

$$\mu_x = E[X] = \begin{cases} \sum_i x_i f_x(x_i) & \text{διακριτή} \\ \int_{-\infty}^{+\infty} x f_x(x) dx & \text{συνεχής} \end{cases}$$

Εμφάνως, η μ_x δίνει το τύπο μέσο της
θεωρητικής συνάρτησης πυκνότητας πιθανότητας

ⓑ Διασπορά

$$\text{Var}[X] = E[(X - \mu_X)^2] =$$

$$\sum_i (x_i - \mu_X)^2 f_X(x_i) \text{ (για } x_i \text{)} \quad \text{ή}$$

$$\int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx \text{ (για } x \text{)} \quad \text{ή}$$



$\text{Var}[X] < \text{Var}[Y]$

(ή το ποσό του ανοιγματος της διασποράς)

↳ τυπική απόκλιση $= \sqrt{\text{Var}[X]} =$ ιδίως ποσός με τυχαία μεταβολή

① Ασυμμετρία.

$$a_x = E[(x - \mu_x)^3]$$

↓

συντελεστής ασυμμετρίας $\gamma_x = \frac{a_x}{\sigma_x^3}$

a_x = μέτρο της ασυμμετρίας.



Kurtosis

$$\alpha_x = E[(\bar{X} - \mu_x)^4]$$

↓

$$K_x = \frac{\alpha_x}{\sigma_x^4} = \frac{\text{fourth moment}}{\text{variance}^2}$$

Πίνακας 5.1. Συνήθεις Κατανομές και οι Χαρακτηρίζουσες Παράμετροι

Κατανομή	Συνάρτηση Πυκνότητας Πιθανότητας (ΣΠΠ) ή Μέγας Πιθανότητας (ΣΜΠ)	Παράμετροι	Σχέση με Μέση τιμή και Διασπορά
Διωνομική	$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, x=0, 1, 2, \dots, n$	p	$E(X) = np$ $Var(X) = np(1-p)$
Γεωμετρική	$p_X(x) = p(1-p)^{x-1}, x=0, 1, 2, \dots$	p	$E(X) = 1/p$ $Var(X) = (1-p)/p^2$
Poisson	$p_X(x) = \frac{(vt)^x}{x!} e^{-vt}, x=0, 1, 2, \dots$	v	$E(X) = vt$ $Var(X) = vt$
Εκθετική	$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, x \geq 0$	λ	$E(X) = \lambda$ $Var(X) = \lambda^2$
Γάμμα	$f_X(x) = \frac{v(vx)^{k-1} e^{-vx}}{\Gamma(k)}, x \geq 0$	v, k	$E(X) = k/v$ $Var(X) = k/v^2$
Κανονική	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], -\infty < x < \infty$	μ, σ	$E(X) = \mu$ $Var(X) = \sigma^2$
Λογαριθμική Κανονική	$f_X(x) = \frac{1}{\sqrt{2\pi}\zeta x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right], x > 0$	λ, ζ	$E(X) = \exp(\lambda + \frac{1}{2}\zeta^2)$ $Var(X) = E^2(X)[e^{\zeta^2} - 1]$
Rayleigh	$f_X(x) = \frac{x}{a^2} \exp\left[-\frac{1}{2}\left(\frac{x}{a}\right)^2\right], x \geq 0$	a	$E(X) = \sqrt{\frac{\pi}{2}}a$ $Var(X) = \left(2 - \frac{\pi}{2}\right)a^2$
Ομοιόμορφη	$f_X(x) = \frac{1}{b-a}, a < x < b$	a, b	$E(X) = (a+b)/2$ $Var(X) = \frac{1}{12}(b-a)^2$
Τριγωνική	$f_X(x) = \frac{2}{b-a}\left(\frac{x-a}{u-a}\right), a \leq x \leq u$ $= \frac{2}{b-a}\left(\frac{b-x}{b-u}\right), u \leq x \leq b$	a, b, u	$E(X) = \frac{1}{3}(a+b+u)$ $Var(X) = \frac{1}{18}(a^2+b^2+u^2 - ab - au - bu)$
Βήτα	$f_X(x) = \frac{1}{B(q,r)} \frac{(x-a)^{q-1}(b-x)^{r-1}}{(b-a)^{q+r-1}}, a \leq x \leq b$	a, b, q, r	$E(X) = a + \frac{q}{q+r}(b-a)$ $Var(X) = \frac{qr}{(q+r)^2(q+r+1)}(b-a)^2$

Οι παραπάνω
ροπές
προσδιορίζονται
από τη θεωρητική
κατανομή
πιθανότητας ως
συνάρτηση των
παραμέτρων της
κατανομής και...

Κανονική κατανομή

- Διάμεσος= μέσο όρο!!!

Σύγκριση αμερόληπτων εκτιμητών και ροπών

- **Εξίσωση**
- των αμερόληπτων εκτιμητών μέσου όρου, τυπικής απόκλισης, συντελεστών κύρτωσης και ασυμμετρίας **του δείγματος,**
- με αυτές όπως προκύπτουν **με τη μέθοδο των ροπών από τις θεωρητικές κατανομές πιθανότητας** (βλπ. αμέσως προηγούμενες διαφάνειες).

Ανάλυση παράγοντα συχνότητας

Παράγοντας Συχνότητας

- Αρχικά επιλέγεται κατανομή
- Ο Chow πρότεινε την **«αντίστροφη πορεία»** από την πιθανότητα να προσδιοριστεί το κατώφλι της τυχαίας μεταβλητής x_T

$$x_T = \bar{x} + K_T s$$

- όπου

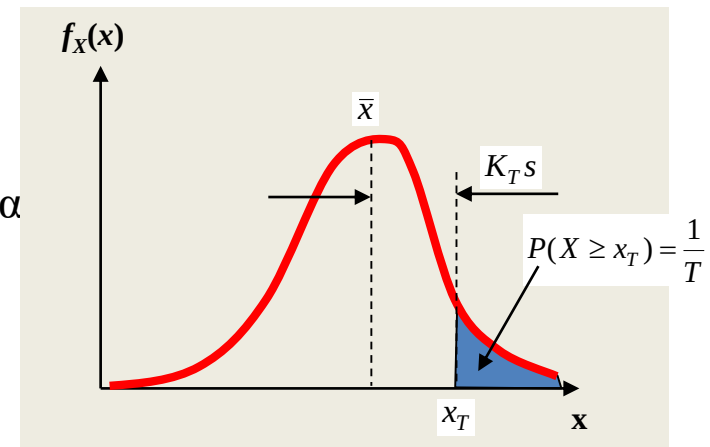
x_T = Εκτιμώμενο κατώφλι

K_T = π.σ με βάση την πιθανότητα και την θεωρητική κα

T = περίοδος επαναφοράς

$\bar{x}$ = μέσος όρος δείγματος

s = τυπική απόκλιση δείγματος



Κανονική κατανομή

- Πυκνότητα πιθανότητας

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

$$K_T = \frac{x_T - \bar{x}}{s} = z_T$$

- $Z=K$

$$x_T = \bar{x} + K_T s = \bar{x} + z_T s$$

- Π.Χ.: Περίοδος επαναφοράς μεγίστων 50 χρόνων

$$T = 50; p = \frac{1}{50} = 0.02 \rightarrow P(Q \geq q) = \frac{1}{50} \rightarrow P(Q \leq q) = 1 - \frac{1}{50} = 0.98$$

$$\rightarrow K_{50} = z_{50} = 2.054$$

Από την Εξ. 2.57 προκύπτει ότι

$$K = \frac{x - \bar{x}}{s} \quad \text{άρα} \quad K = z$$

2.4.2 Κατανομή μεγίστων ακραίων τιμών

$$y = \frac{a+x}{c} \quad a = \gamma c - \mu \quad c = \frac{\sqrt{6}}{\pi} \sigma \quad (2.61)$$

$$x = \mu + \sigma \left[-\frac{\sqrt{6}}{\pi} (\gamma - y) \right] \quad (2.62)$$

$$y = -\ln \ln \left(\frac{T}{T-1} \right) \quad (2.63)$$

$$K = -\frac{\sqrt{6}}{\pi} \left[\gamma + \ln \ln \left(\frac{T}{T-1} \right) \right] \quad N \rightarrow \infty \quad (2.64)$$

για N πεπερασμένο από την εκτίμηση των παραμέτρων a, c προκύπτει :

$$a_n = c_n y_n - \bar{x} \quad c_n = \frac{s}{\sigma_n} \quad (2.65)$$

$$x = \bar{x} + s \left[\frac{y - \bar{y}_n}{\sigma_n} \right] \quad K = -\frac{\ln \ln \left(\frac{T}{T-1} \right) + y_n}{\sigma_n} \quad (2.66)$$

Έλεγχος αν η θεωρητική κατανομή πιθανότητας που επιλέχθηκε είναι κατάλληλη

Τεστ καταλληλότητας

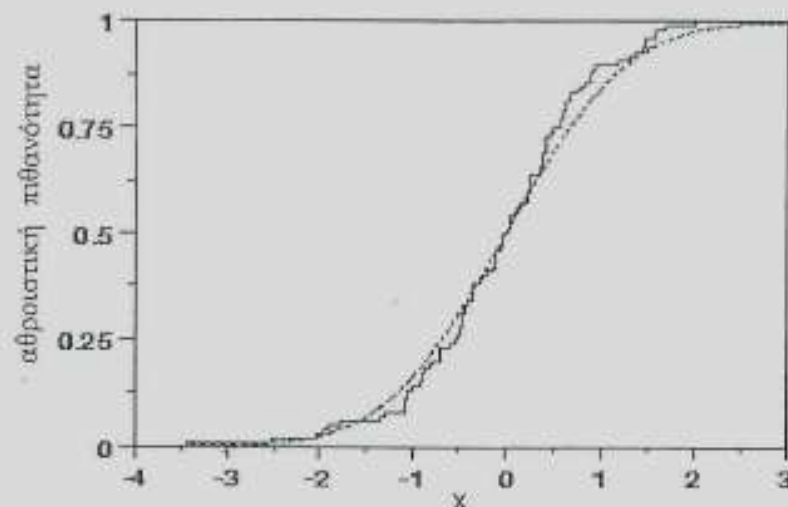
2.6.2 Κριτήριο Kolmogorov – Smirnov

Το κριτήριο αυτό στηρίζεται στην εμπειρική συνάρτηση πιθανότητας E_N που ορίζεται ως εξής:

$$E_N = \sum N_i / N \quad (2.78)$$

όπου N ο αριθμός των παρατηρήσεων $x_1, x_2, \dots, x_N$ και N_i είναι ο αριθμός παρατηρήσεων μικρότερος του x_i όταν τα x_i είναι διαταγμένα κατά αύξουσα σειρά.

Στο Σχήμα 2.19 φαίνεται η γραφική παράσταση της σύγκρισης μιας εμπειρικής κατανομής και της κανονικής αθροιστικής κατανομής για 100 τυχαίους αριθμούς. Το κριτήριο Kolmogorov – Smirnov στηρίζεται στη μέγιστη απόκλιση των δύο κατανομών.



Σχ. 2.16 Εμπειρική κατανομή και Κανονική αθροιστική κατανομή

Το κριτήριο αυτό μαθηματικά εκφράζεται ως εξής :

$$D = \max_{x_i} \left| F(x_i) - \sum N_i / N \right| \quad (2.79)$$

όπου η $F(x_i)$ η θεωρητική αθροιστική κατανομή της κατανομής που

Πιν. 2.10: Κρίσιμες τιμές D_{cr} του ελέγχου Kolmogorov-Smirnov.

Μέγεθος δείγματος N	α				
	0.20	0.15	0.10	0.05	0.01
1	0.90	0.93	0.95	0.98	0.99
2	0.68	0.73	0.78	0.84	0.93
3	0.57	0.60	0.64	0.71	0.83
4	0.49	0.53	0.56	0.62	0.73
5	0.45	0.47	0.51	0.56	0.67
6	0.41	0.44	0.47	0.52	0.62
7	0.38	0.41	0.44	0.49	0.58
8	0.36	0.38	0.41	0.46	0.54
9	0.34	0.36	0.39	0.43	0.51
10	0.32	0.34	0.37	0.41	0.49
11	0.31	0.33	0.35	0.39	0.47
12	0.30	0.31	0.34	0.38	0.45
13	0.28	0.30	0.33	0.36	0.43
14	0.27	0.29	0.31	0.35	0.42
15	0.27	0.28	0.30	0.34	0.40
16	0.26	0.27	0.30	0.33	0.39
17	0.25	0.27	0.29	0.32	0.38
18	0.24	0.26	0.28	0.31	0.37
19	0.24	0.25	0.27	0.30	0.36
20	0.23	0.25	0.26	0.29	0.35
25	0.21	0.22	0.24	0.26	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.21	0.23	0.27
40	0.17	0.18	0.19	0.21	0.25
45	0.16	0.17	0.18	0.20	0.24
50	0.15	0.16	0.17	0.19	0.23
$N > 50$	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$