

ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΣΤΑΤΙΣΤΙΚΗ ΤΩΝ ΕΠΙΧΕΙΡΗΣΕΩΝ

Κεφάλαιο 7

Τυχαίες Μεταβλητές και

Διακριτές Κατανομές Πιθανοτήτων

- *Επιμέλεια παρουσιάσεων: Δρ. Αλέκα Καλαπόδη*

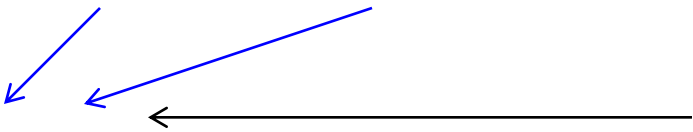
Τυχαίες Μεταβλητές ...

Μια *τυχαία μεταβλητή* είναι μια συνάρτηση ή ένας κανόνας που αντιστοιχίζει έναν αριθμό σε κάθε αποτέλεσμα ενός πειράματος.

Με διαφορετικά λόγια, η *τιμή* της τυχαίας μεταβλητής είναι ένα αριθμητικό ενδεχόμενο.

Αντί να βλέπουμε τα αποτελέσματα της ρίψης ενός νομίσματος ως

{κορώνα, γράμματα} σκεφτείτε τα ως


{1, 0} “πόσες φορές ήρθε κορώνα”
(αριθμητικά ενδεχόμενα)

Δύο Τύποι Τυχαίων Μεταβλητών ...

Διακριτή Τυχαία Μεταβλητή

- αυτή που παίρνει *πεπερασμένο* πλήθος τιμών
- π.χ. τιμές ρίψης δύο ζαριών: 2, 3, 4, ..., 12

Συνεχής Τυχαία Μεταβλητή

- αυτή που οι τιμές της *δεν είναι διακριτές*
- π.χ. χρόνος (30.1 λεπτά; 30.10000001 λεπτά;)

Αναλογία:

Οι ακέραιοι είναι Διακριτοί, ενώ οι Πραγματικοί αριθμοί είναι συνεχείς.

Κατανομές Πιθανοτήτων ...

Μια *κατανομή πιθανοτήτων* είναι ένας πίνακας, ένας τύπος, ή γράφημα που περιγράφει τις τιμές μιας τυχαίας μεταβλητής και την πιθανότητα κάθε τιμής.

Αφού περιγράφουμε μια **τυχαία μεταβλητή** (η οποία μπορεί να είναι διακριτή ή συνεχής) έχουμε δύο τύπους κατανομής πιθανοτήτων:

- Διακριτές Κατανομές Πιθανοτήτων, και
- Συνεχείς Κατανομές Πιθανοτήτων

Συμβολισμός ...

Ένα κεφαλαίο γράμμα αναπαριστά το *όνομα* της τυχαίας μεταβλητής, συνήθως **X**.

Το αντίστοιχο μικρό γράμμα αναπαριστά την *τιμή* της τυχαίας μεταβλητής.

Η πιθανότητα η τυχαία μεταβλητή **X** να ισούται με x είναι:

$$P(\mathbf{X} = x)$$

ή πιο απλά

$$P(x)$$

Διακριτές Κατανομές Πιθανοτήτων ...

Οι πιθανότητες των τιμών μιας *διακριτής τυχαίας μεταβλητής* μπορούν να υπολογιστούν με τη βοήθεια εργαλείων όπως το δέντρο πιθανοτήτων ή εφαρμόζοντας τον ορισμό της πιθανότητας, με προϋποθέσεις:

$$1. 0 \leq P(x) \leq 1 \text{ για κάθε } x$$

$$2. \sum_{\text{όλων των } x_i} P(x) = 1$$

Παράδειγμα 7.1

Η Στατιστική σύνοψη των ΗΠΑ δημοσιεύεται κάθε χρόνο. Περιέχει μεγάλη ποικιλία πληροφοριών που βασίζονται στην απογραφή και σε άλλες πηγές. Στόχος είναι η παροχή πληροφοριών για διάφορα χαρακτηριστικά της ζωής των κατοίκων. Μια από τις ερωτήσεις ήταν το πλήθος των έγχρωμων τηλεοράσεων ανά νοικοκυριό. Τα δεδομένα βρίσκονται στον ακόλουθο πίνακα. Να κατασκευαστεί η κατανομή πιθανοτήτων της τυχαίας μεταβλητής που εκφράζει το πλήθος των έγχρωμων τηλεοράσεων ανά νοικοκυριό.

Παράδειγμα 7.1

<u>Πλήθος έγχρωμων τηλεοράσεων</u>	<u>Πλήθος νοικοκυριών (σε χιλιάδες)</u>
0	1,218
1	32,379
2	37,961
3	19,387
4	7,714
5	2,842
<hr/>	
Σύνολο	101,501

Παράδειγμα 7.1

Η κατανομή πιθανοτήτων μπορεί να υπολογιστεί από τις σχετικές συχνότητες.

<i># των τηλεοράσεων</i>	<i># των νοικοκυριών</i>	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		1.000

$$1,218 \div 101,501 = 0.012$$

Π.χ. $P(\mathbf{X}=4) = P(4) = 0.076 = 7.6\%$

Παράδειγμα 7.1

Π.χ. ποια είναι η πιθανότητα να υπάρχει τουλάχιστον μία τηλεόραση αλλά όχι περισσότερες από τρεις σε ένα σπίτι;

# των τηλεοράσεων	# των νοικοκυριών	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		1.000

“τουλάχιστον μία τηλεόραση αλλά όχι περισσότερες από τρεις”

$$P(1 \leq \mathbf{X} \leq 3) = P(1) + P(2) + P(3) = .319 + .374 + .191 = .884$$

Παράδειγμα 7.2...

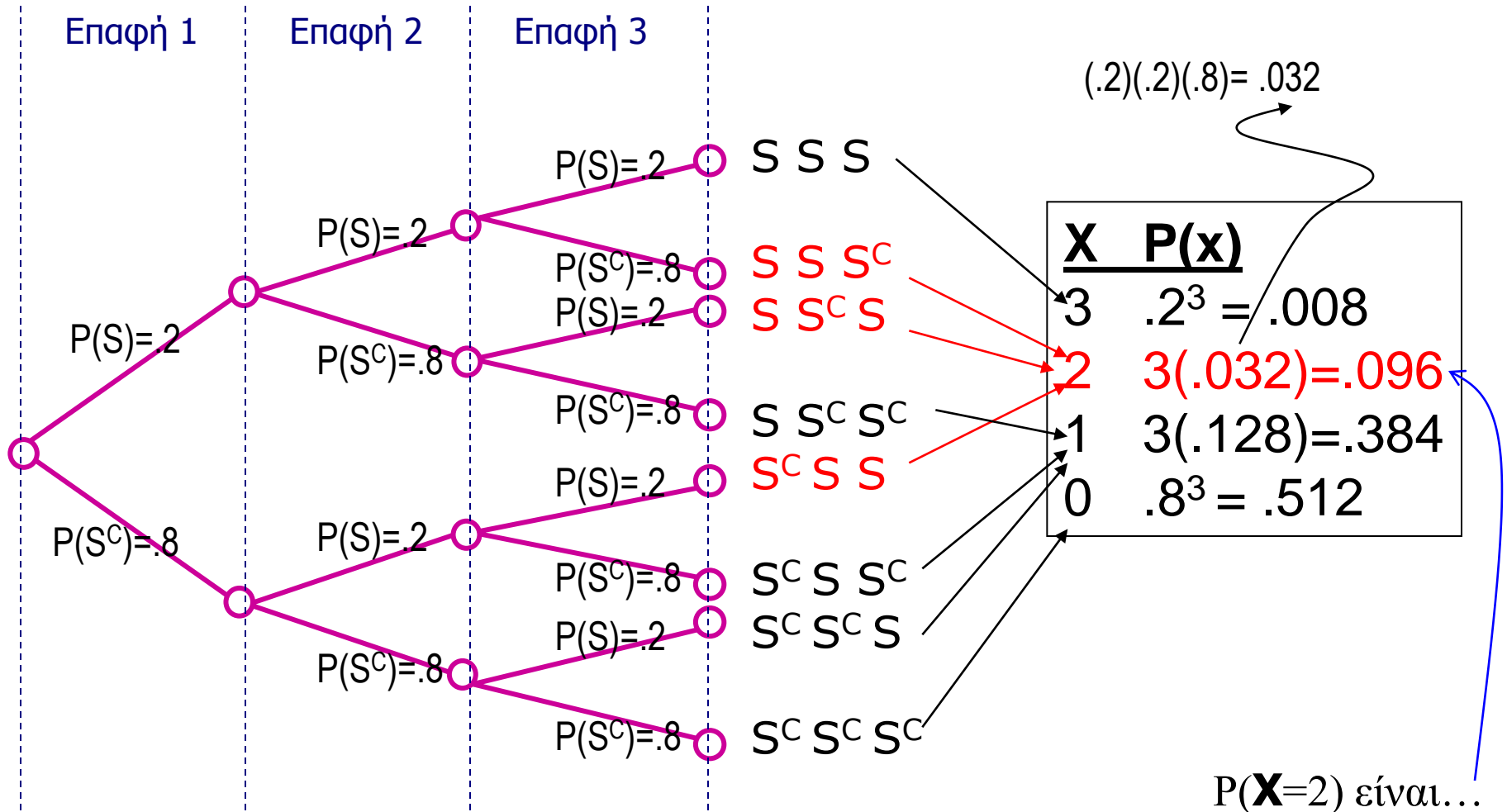
Ένας πωλητής αμοιβαίων κεφαλαίων έχει κανονίσει να μιλήσει με τρία άτομα αύριο. Βασιζόμενος στην εμπειρία του γνωρίζει ότι υπάρχει σε κάθε επαφή 20% πιθανότητα να πετύχει πώληση. Να καθοριστεί η κατανομή πιθανοτήτων του αριθμού των πωλήσεων που θα κάνει ο πωλητής.

Έστω S η επιτυχία, δηλ. να πετύχει πώληση $P(S)=.20$

Τότε S^C είναι να μην πετύχει πώληση, και $P(S^C)=.80$

Παράδειγμα 7.2...

Κατασκευάζοντας την Κατανομή Πιθανοτήτων ...



P(**X**=2) είναι...

Πληθυσμοί και Κατανομές Πιθανοτήτων ...

Η διακριτή κατανομή πιθανοτήτων αναπαριστά έναν **πληθυσμό**

Παράδειγμα 7.1 το πλήθος των τηλεοράσεων ανά νοικοκυριό

Παράδειγμα 7.2 το πλήθος των πωλήσεων

Αφού έχουμε **πληθυσμούς**, μπορούμε να τους περιγράψουμε υπολογίζοντας διάφορες **παραμέτρους**.

Π.χ. τον μέσο και τη διακύμανση του πληθυσμού.

Μέσος Πληθυσμού (Αναμενόμενη Τιμή)

Ο μέσος πληθυσμού είναι ο *σταθμισμένος μέσος όρος* όλων των τιμών του. Οι συντελεστές στάθμισης είναι οι πιθανότητες.

Η παράμετρος αυτή καλείται και αναμενόμενη τιμή της X και συμβολίζεται με $E(X)$.

$$E(X) = \mu = \sum_{\text{όλων των } x} xP(x)$$

Διασπορά Πληθυσμού ...

Η διασπορά πληθυσμού υπολογίζεται με παρόμοιο τρόπο. Είναι ο σταθμισμένος αριθμητικός μέσος των *τετραγώνων των αποκλίσεων* από το μέσο.

$$V(X) = \sigma^2 = \sum_{\text{όλων των } x} (x - \mu)^2 P(x)$$

Όπως πριν, υπάρχει και “σύντομος” τύπος ...

$$V(X) = \sigma^2 = \sum_{\text{όλων των } x} x^2 P(x) - \mu^2$$

Η τυπική απόκλιση είναι πάλι:

$$\sigma = \sqrt{\sigma^2}$$

Παράδειγμα 7.3...

Βρείτε τον **μέσο**, τη διασπορά και την τυπική απόκλιση του πληθυσμού που εκφράζει το πλήθος των έγχρωμων τηλεοράσεων ανά νοικοκυριό ... (από το Παράδειγμα 7.1)

<i># των τηλεοράσεων</i>	<i># των νοικοκυριών</i>	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		1.000

$$E(X) = \mu = \sum_{\text{all } x} xP(x) = 0 \cdot P(0) + 1 \cdot P(1) + \dots + 5 \cdot P(5)$$

$$= 0(.012) + 1(.319) + 2(.374) + 3(.191) + 4(.076) + 5(.028)$$

$$= \mathbf{2.084}$$

Παράδειγμα 7.3...

Βρείτε τον μέσο, τη **διασπορά** και την τυπική απόκλιση του πληθυσμού που εκφράζει το πλήθος των έγχρωμων τηλεοράσεων ανά νοικοκυριό ... (από το Παράδειγμα 7.1)

<i># των τηλεοράσεων</i>	<i># των νοικοκυριών</i>	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		1.000

$$V(X) = \sigma^2 = \sum_{\text{all } x} (x - \mu)^2 P(x)$$

$$= (0 - 2.084)^2(.012) + (1 - 2.084)^2(.319) + \dots + (5 - 2.084)^2(.028)$$
$$= \mathbf{1.107}$$

Παράδειγμα 7.3...

Βρείτε τον μέσο, τη διασπορά και την **τυπική απόκλιση** του πληθυσμού που εκφράζει το πλήθος των έγχρωμων τηλεοράσεων ανά νοικοκυριό ... (από το Παράδειγμα 7.1)

<i># των τηλεοράσεων</i>	<i># των νοικοκυριών</i>	x	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
	<u>101,501</u>		1.000

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.107} = \mathbf{1.052}$$

Νόμοι της Αναμενόμενης Τιμής ...

$$E(c) = c$$

Η αναμενόμενη τιμή μιας σταθεράς (c) είναι απλά η τιμή της σταθεράς.

$$E(X + c) = E(X) + c$$

$$E(cX) = cE(X)$$

Μπορούμε να “βγάλουμε” μια σταθερά έξω από την έκφραση της αναμενόμενης τιμής (είτε ως προσθετέο μιας τυχαίας μεταβλητής X είτε ως συντελεστή μια τυχαίας μεταβλητής X).

Παράδειγμα 7.4...

Οι μηνιαίες πωλήσεις έχουν μέσο \$25.000 και τυπική απόκλιση \$4.000. Τα κέρδη υπολογίζονται πολλαπλασιάζοντας τις πωλήσεις με 30% και αφαιρώντας το σταθερό κόστος των \$6.000.

Να βρείτε το μέσο μηνιαίο κέρδος.

1) Περιγράφουμε το πρόβλημα με αλγεβρικούς όρους:

οι πωλήσεις έχουν μέσο \$25.000 → $E(\text{Πωλήσεις}) = 25.000$

τα κέρδη υπολογίζονται από... → $\text{Κέρδη} = .30(\text{Πωλήσεις}) - 6.000$

Παράδειγμα 7.4...

Οι μηνιαίες πωλήσεις έχουν μέσο \$25.000 και τυπική απόκλιση \$4.000. Τα κέρδη υπολογίζονται πολλαπλασιάζοντας τις πωλήσεις με 30% και αφαιρώντας το σταθερό κόστος των \$6.000.

*Να βρείτε το **μέσο** μηνιαίο κέρδος.*

$$\begin{aligned} E(\text{Κέρδος}) &= E[.30(\text{Πωλήσεις}) - 6.000] \\ &= E[.30(\text{Πωλήσεις})] - 6.000 && \text{[κανόνας \#2]} \\ &= .30E(\text{Πωλήσεις}) - 6.000 && \text{[κανόνας \#3]} \\ &= .30(25.000) - 6.000 = 1.500 \end{aligned}$$

Άρα, το μέσο μηνιαίο κέρδος είναι **\$1.500**

Νόμοι Διασποράς ...

$$V(c) = 0$$

Η διασπορά μιας σταθεράς (c) είναι μηδέν.

$$V(X + c) = V(X)$$

Η διασπορά μιας τυχαίας μεταβλητής και μιας σταθεράς είναι απλά η διασπορά της τυχαίας μεταβλητής (από το προηγούμενο).

$$V(cX) = c^2V(X)$$

Η διασπορά μιας τυχαίας μεταβλητής επί έναν σταθερό συντελεστή είναι το τετράγωνο του συντελεστή επί τη διασπορά της τυχαίας μεταβλητής.

Παράδειγμα 7.4...

Οι μηνιαίες πωλήσεις έχουν μέσο \$25.000 και τυπική απόκλιση \$4.000. Τα κέρδη υπολογίζονται πολλαπλασιάζοντας τις πωλήσεις με 30% και αφαιρώντας το σταθερό κόστος των \$6.000.

*Να βρείτε την **τυπική απόκλιση** των μηνιαίων κερδών.*

1) Περιγράφουμε το πρόβλημα με αλγεβρικούς όρους :

οι πωλήσεις έχουν τυπική απόκλιση \$4.000

$$\rightarrow V(\text{Πωλήσεις}) = 4.000^2 = 16.000.000$$

(θυμηθείτε τη σχέση $\sigma = \sqrt{\sigma^2}$ μεταξύ τυπικής απόκλισης και διακύμανσης)

τα κέρδη υπολογίζονται από... \rightarrow Κέρδη = .30(Πωλήσεις) – 6.000

Παράδειγμα 7.4...

Οι μηνιαίες πωλήσεις έχουν μέσο \$25.000 και τυπική απόκλιση \$4.000. Τα κέρδη υπολογίζονται πολλαπλασιάζοντας τις πωλήσεις με 30% και αφαιρώντας το σταθερό κόστος των \$6.000.

*Να βρείτε την **τυπική απόκλιση** των μηνιαίων κερδών.*

2) Η **διασπορά** των κερδών είναι = $V(\text{Κέρδος})$

$$=V[.30(\text{Πωλήσεις}) - 6.000]$$

$$=V[.30(\text{Πωλήσεις})] \quad [\text{κανόνας \#2}]$$

$$=(.30)^2V(\text{Πωλήσεις}) \quad [\text{κανόνας \#3}]$$

$$=(.30)^2(16.000.000) = 1.440.000$$

Η **τυπική απόκλιση** είναι η τετραγωνική ρίζα της **διασποράς**,
άρα η τυπική απόκλιση του Κέρδους = $(1.440.000)^{1/2} = \mathbf{\$1.200}$

Παράδειγμα 7.4 (σύνοψη)

Οι μηνιαίες πωλήσεις έχουν μέσο \$25.000 και τυπική απόκλιση \$4.000. Τα κέρδη υπολογίζονται πολλαπλασιάζοντας τις πωλήσεις με 30% και αφαιρώντας το σταθερό κόστος των \$6.000.

Βρείτε το μέσο και την τυπική απόκλιση των μηνιαίων κερδών

Το **μέσο** μηνιαίο κέρδος είναι **\$1.500**

Η **τυπική απόκλιση** των μηνιαίων κερδών είναι **\$1.200**

Διωνυμική Κατανομή ...

Η *διωνυμική κατανομή* είναι η κατανομή πιθανοτήτων που προκύπτει από ένα “*διωνυμικό πείραμα*”. Τα διωνυμικά πειράματα έχουν τις ακόλουθες ιδιότητες:

Καθορισμένο αριθμό δοκιμών, που συμβολίζεται με n .

Κάθε δοκιμή έχει δύο δυνατά αποτελέσματα, “επιτυχία” και “αποτυχία”.

$$P(\text{επιτυχία})=p$$

(και επομένως: $P(\text{αποτυχία})=1-p$), για κάθε δοκιμή.

Οι δοκιμές είναι *ανεξάρτητες*, δηλαδή το αποτέλεσμα μιας δοκιμής δεν επηρεάζει το αποτέλεσμα οποιασδήποτε άλλης.

Επιτυχία και Αποτυχία ...

...είναι απλά ταμπέλες για ένα διωνυμικό πείραμα, δεν σημαίνουν κάτι.

Για παράδειγμα η ρίψη ενός νομίσματος έχει ως αποτέλεσμα κορώνα ή γράμματα. Εάν ορίσουμε ως επιτυχία το να έρθει “κορώνα” τότε υποχρεωτικά το να έρθει “γράμματα” θεωρείται αποτυχία.

Άλλα παραδείγματα:

Ένας υποψήφιος κερδίζει ή χάνει

Ένας υπάλληλος είναι άντρας ή γυναίκα

Διωνυμική Τυχαία Μεταβλητή ...

Η τυχαία μεταβλητή ενός διωνυμικού πειράματος ορίζεται ως το πλήθος των επιτυχιών σε n δοκιμές, και καλείται *διωνυμική τυχαία μεταβλητή*.

Π.χ. ρίχνουμε ένα νόμισμα 10 φορές...

- 1) Καθορισμένο πλήθος δοκιμών ✓ $n=10$
- 2) Κάθε δοκιμή έχει δύο πιθανά αποτελέσματα ✓ {κορώνα (επιτυχία), γράμματα (αποτυχία)}
- 3) $P(\text{επιτυχία})=0.50$; $P(\text{αποτυχία})=1-0.50=0.50$ ✓
- 4) Οι δοκιμές είναι ανεξάρτητες ✓ (δηλ. το αποτέλεσμα να έρθει κορώνα στην πρώτη ρίψη δεν θα έχει επίδραση στις επόμενες ρίψεις).

Επομένως η ρίψη ενός νομίσματος δέκα φορές είναι διωνυμικό πείραμα.

Διωνυμική Τυχαία Μεταβλητή ...

Η διωνυμική τυχαία μεταβλητή *μετράει* το πλήθος των επιτυχιών σε n δοκιμές του διωνυμικού πειράματος. Παίρνει τιμές $0, 1, 2, \dots, n$. Άρα, είναι μια *διακριτή* τυχαία μεταβλητή.

Για να υπολογίσουμε την πιθανότητα που αντιστοιχεί σε κάθε τιμή χρησιμοποιούμε συνδυαστική:

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{για } x=0, 1, 2, \dots, n$$

Αδιάβαστος φοιτητής ...

Ένας φοιτητής στατιστικής είναι αδιάβαστος και θα απαντήσει στην τύχη στις εξετάσεις. Η εξέταση έχει 10 ερωτήσεις πολλαπλού τύπου. Κάθε ερώτηση έχει πέντε πιθανές απαντήσεις, με μόνο μία σωστή. Ο φοιτητής ελπίζει να πετύχει τη σωστή απάντηση σε κάθε ερώτηση.

Ποια είναι η πιθανότητα να μην πετύχει καμία σωστή απάντηση;

Ποια είναι η πιθανότητα να πετύχει δύο σωστές απαντήσεις;

Αδιάβαστος φοιτητής ...

Ένας φοιτητής στατιστικής είναι αδιάβαστος και θα απαντήσει στην τύχη στις εξετάσεις. Η εξέταση έχει **10 ερωτήσεις πολλαπλού τύπου**. Κάθε ερώτηση έχει **πέντε πιθανές απαντήσεις, με μόνο μία σωστή**. Ο φοιτητής ελπίζει να πετύχει τη σωστή απάντηση σε κάθε ερώτηση.

Αλγεβρικά: **$n=10$** , και **$P(\text{επιτυχία}) = 1/5 = .20$**

Αδιάβαστος φοιτητής ...

Είναι διωνυμικό πείραμα; Ελέγχουμε:

✓ Υπάρχει καθορισμένο πλήθος δοκιμών ($n=10$).

✓ Κάθε απάντηση είναι είτε σωστή είτε λάθος.

Η πιθανότητα σωστής απάντησης είναι ($P(\text{επιτυχία})=.20$) δεν αλλάζει από ερώτηση σε ερώτηση.

✓ Κάθε απάντηση είναι ανεξάρτητη από τις άλλες.

Αδιάβαστος φοιτητής ...

$$n=10, \text{ και } P(\text{επιτυχία}) = .20$$

Ποια είναι η πιθανότητα να μην πετύχει *καμία σωστή απάντηση*;

Δηλ. # επιτυχιών, $x = 0$, επομένως θέλουμε την $P(x=0)$


$$\begin{aligned} P(0) &= \frac{10!}{0!(10-0)!} (.2)^0 (1-.2)^{10-0} \\ &= 1(1)(.8)^{10} = .1074 \end{aligned}$$

Έχει περίπου 11% πιθανότητα να μην πετύχει καμία σωστή απάντηση

Αδιάβαστος φοιτητής ...

n=10, και **P(επιτυχία) = .20**

Ποια είναι η πιθανότητα να πετύχει *δύο σωστές* απαντήσεις;
Δηλ. # επιτυχιών, $x = 2$, άρα θέλουμε την $P(x=2)$

$$\begin{aligned} P(2) &= \frac{10!}{2!(10-2)!} (.2)^2 (1-.2)^{10-2} \\ &= 45(.04)(.1678) = .3020 \end{aligned}$$


Έχει περίπου 30% πιθανότητα να πετύχει ακριβώς δύο σωστές απαντήσεις

Αθροιστική Πιθανότητα ...

Μέχρι εδώ, χρησιμοποιήσαμε την διωνυμική κατανομή πιθανότητας για να βρούμε πιθανότητες συγκεκριμένων τιμών x . Για να απαντήσουμε την ερώτηση:

“Βρείτε την πιθανότητα αποτυχίας στις εξετάσεις”

Απαιτείται η *αθροιστική πιθανότητα*, δηλαδή, $P(X \leq x)$

Εάν ο βαθμός είναι μικρότερος από 50% (δηλ. 5 ερωτήσεις από τις 10), τότε έχουμε αποτυχία.


Άρα, θέλουμε την $P(X \leq 4)$ για να απαντήσουμε

Αδιάβαστος φοιτητής ...

$$P(X \leq 4) = P(0) + P(1) + P(2) + P(3) + P(4)$$

Ήδη ξέρουμε ότι $P(0) = .1074$ και $P(2) = .3020$.
Χρησιμοποιούμε τον τύπο και έχουμε:

$$P(1) = .2684, P(3) = .2013, \text{ και } P(4) = .0881$$


$$\text{Έχουμε } P(X \leq 4) = .1074 + .2684 + \dots + .0881 = \mathbf{.9672}$$


*Άρα, είναι περίπου 97% πιθανό να αποτύχει στις εξετάσεις
εάν απαντήσει στην τύχη ...*

Διωνυμικός Πίνακας ...

Είναι δύσκολο να υπολογίζουμε με το χέρι τις διωνυμικές πιθανότητες. Υπάρχει πιο εύκολος τρόπος, ο [Πίνακας 1](#) στο Παράρτημα Β. Για το παράδειγμα του φοιτητή, **n=10**, άρα πρώτα πρέπει να κοιτάξουμε το σωστό πίνακα!

n = 10



k	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

Διωνυμικός Πίνακας ...

Οι πιθανότητες που υπάρχουν είναι *αθροιστικές*,
δηλ. $P(X \leq k)$ – k είναι ο δείκτης γραμμής, οι στήλες
ορίζονται από την $P(\text{επιτυχία}) = p$

$n = 10$

k	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

Διωνυμικός Πίνακας ...

“Ποια είναι η πιθανότητα να μην πετύχει *καμία σωστή* απάντηση;”

δηλ. ποια είναι η $P(X = 0)$, εάν $P(\text{επιτυχία}) = .20$ και $n=10$;

n = 10		0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
k	0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
	3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
	4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
	5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
	6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

$$P(X = 0) = P(X \leq 0) = .1074$$

Διωνυμικός Πίνακας ...

“Ποια είναι η πιθανότητα να πετύχει *δύο σωστές* απαντήσεις;”
δηλ. ποια είναι η $P(X = 2)$, εάν $P(\text{επιτυχία}) = .20$ και $n=10$;

n = 10		0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
k	0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
	3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
	4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
	5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
	6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

$$P(X = 2) = P(X \leq 2) - P(X \leq 1) = .6778 - .3758 = .3020$$

θυμηθείτε, ο πίνακας δίνει *αθροιστικές* πιθανότητες...

Διωνυμική Κατανομή ...

Βρείτε την πιθανότητα αποτυχίας στις εξετάσεις

δηλ. ποια είναι η $P(X \leq 4)$, εάν $P(\text{επιτυχία}) = .20$ και $n=10$;

n = 10		0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
0	k	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1		0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2		0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3		1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4		1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5		1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6		1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7		1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8		1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

$$P(X \leq 4) = .9672$$

Διωνυμικός Πίνακας ...

Ο διωνυμικός πίνακας δίνει αθροιστικές πιθανότητες για $P(X \leq k)$, αλλά όπως είδαμε,

$$P(X = k) = P(X \leq k) - P(X \leq [k-1])$$

Αντίστοιχα, για πιθανότητες $P(X \geq k)$, έχουμε:

$$P(X \geq k) = 1 - P(X \leq [k-1])$$

Διωνυμική Κατανομή ...

Όπως μάλλον περιμένετε, οι στατιστικοί έχουν αναπτύξει γενικούς τύπους για το μέσο, τη διασπορά και την τυπική απόκλιση μιας διωνυμικής τυχαίας μεταβλητής:

$$\mu = np$$

$$\sigma^2 = np(1 - p)$$

$$\sigma = \sqrt{np(1 - p)}$$

Κατανομή Poisson ...

Ονομαζόμενη έτσι από τον Simeon Poisson, η *κατανομή Poisson* είναι μια διακριτή κατανομή πιθανοτήτων που αναφέρεται στο πλήθος των γεγονότων (όπως οι επιτυχίες) μέσα σε συγκεκριμένο χρονικό διάστημα ή σε συγκεκριμένο χώρο. Για παράδειγμα:

Ο αριθμός αυτοκινήτων που φτάνουν σε ένα συνεργείο μέσα σε 1 ώρα. (Το χρονικό διάστημα είναι 1 ώρα.)

Ο αριθμός ελαττωμάτων σε ένα τόπι ύφασμα. (Ο συγκεκριμένος χώρος είναι το τόπι υφάσματος.)

Ο αριθμός ατυχημάτων σε 1 ημέρα σε συγκεκριμένο τμήμα του αυτοκινητόδρομου. (Το διάστημα ορίζεται και από το χρόνο και από το χώρο.)

Το Πείραμα Poisson ...

Όπως ένα διωνυμικό πείραμα, ένα *πείραμα Poisson* έχει 4 καθοριστικές ιδιότητες:

Ο αριθμός των επιτυχιών που μπορούν να συμβούν σε ένα διάστημα είναι ανεξάρτητος από τον αριθμό των επιτυχιών σε οποιοδήποτε άλλο διάστημα.

Η πιθανότητα επιτυχίας είναι ίδια για κάθε διάστημα ίσου μήκους.

Η πιθανότητα επιτυχίας είναι ανάλογη του μήκους του διαστήματος.

Η πιθανότητα για πάνω από μία επιτυχία σε ένα διάστημα πλησιάζει το 0 όσο το διάστημα μικραίνει.

Κατανομή Poisson ...

Η *τυχαία μεταβλητή Poisson* είναι το πλήθος των επιτυχιών που συμβαίνουν σε μια χρονική περίοδο ή σε συγκεκριμένο χώρο ενός πειράματος Poisson.

επιτυχίες



Π.χ. Κατά μέσο όρο, **96** φορτηγά φτάνουν στα σύνορα

κάθε ώρα.

Χρονική
περίοδος



Π.χ. Ο αριθμός των τυπογραφικών λαθών σε μια νέα έκδοση έχει μέσο όρο **1.5** ανά **100** **σελίδες.**

επιτυχίες (!)



διάστημα



Κατανομή πιθανοτήτων Poisson ...

Η πιθανότητα μια τυχαία μεταβλητή Poisson να πάρει την τιμή x είναι:

$$P(x) = \frac{e^{-\mu} \mu^x}{x!} \quad \text{για } x = 0, 1, 2, \dots$$

όπου μ είναι το μέσο πλήθος επιτυχιών στο διάστημα

και e είναι η βάση του νεπέριου λογαρίθμου.

Ισχύει: $E(X) = V(X) = \mu$

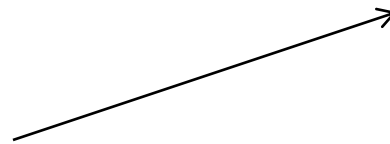
Παράδειγμα 7.12...

Ένας καθηγητής στατιστικής παρατήρησε ότι το πλήθος των τυπογραφικών λαθών στις νέες εκδόσεις διαφέρει σημαντικά από βιβλίο σε βιβλίο. Μετά από ανάλυση, κατέληξε ότι το πλήθος των λαθών ακολουθεί κατανομή Poisson με μέσο 1.5 ανά 100 σελίδες. Ο καθηγητής επιλέγει τυχαία 100 σελίδες ενός βιβλίου. Ποια είναι η πιθανότητα να μην υπάρχουν καθόλου λάθη;

Παράδειγμα 7.12...

Ένας καθηγητής στατιστικής παρατήρησε ότι το πλήθος των τυπογραφικών λαθών στις νέες εκδόσεις διαφέρει σημαντικά από βιβλίο σε βιβλίο. Μετά από ανάλυση, κατέληξε ότι το πλήθος των λαθών ακολουθεί κατανομή Poisson με μέσο 1.5 ανά 100 σελίδες. Ο καθηγητής επιλέγει τυχαία 100 σελίδες ενός βιβλίου. Ποια είναι η πιθανότητα να μην υπάρχουν καθόλου λάθη; Δηλαδή, ποια είναι η $P(X=0)$ με δεδομένο ότι $\mu = 1.5$;

$$P(0) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-1.5} 1.5^0}{0!} = .2231$$



“Υπάρχει περίπου 22% πιθανότητα να μην υπάρχουν λάθη”

Παράδειγμα 7.13...

Αναφερόμαστε στο Παράδειγμα 7.12. Υποθέτουμε ότι ο καθηγητής έχει μόλις λάβει ένα νέο βιβλίο. Παρατηρεί ότι έχει 400 σελίδες.

a Ποια είναι η πιθανότητα να μην υπάρχουν καθόλου λάθη;

b Ποια είναι η πιθανότητα να υπάρχουν πέντε ή λιγότερα λάθη;

Κατανομή Poisson ...

Όπως είπαμε για ένα πείραμα Poisson :

Η πιθανότητα της επιτυχίας είναι ανάλογη με το μέγεθος του διαστήματος

Άρα, γνωρίζοντας το ρυθμό 1.5 λάθος ανά 100 σελίδες, μπορούμε να ορίσουμε μέση τιμή για βιβλίο 400 σελίδων:

$$\mu = 1.5(4) = 6 \text{ λάθη/ 400 σελίδες.}$$

Παράδειγμα 7.13...

Για βιβλίο 400 σελίδων, ποια είναι η πιθανότητα να μην υπάρχουν λάθη;

$$P(X=0) = P(0) = \frac{e^{-\mu} \mu^x}{x!} = \frac{e^{-6} 6^0}{0!} = .002479$$

“υπάρχει πολύ μικρή πιθανότητα να μην υπάρχουν λάθη”

Παράδειγμα 7.13...

Για βιβλίο 400 σελίδων, ποια είναι η πιθανότητα να υπάρχουν **πέντε ή λιγότερα** λάθη;

$$P(X \leq 5) = P(0) + P(1) + \dots + P(5) = \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \dots + \frac{e^{-6}6^5}{5!}$$

Αυτό είναι δύσκολο να υπολογιστεί με το χέρι. Καλύτερα να δούμε τον [Πίνακα 2](#) στο Παράρτημα Β...

... $k=5$, $\mu = 6$, και $P(X \leq k) = .446$

“υπάρχει περίπου 45% πιθανότητα να υπάρχουν 5 ή λιγότερα λάθη”