Πιθανοθεωρητικοί Αλγόριθμοι: Monte Carlo, Ισχύς Ψήφου, Πρώτοι Αριθμοί

Αν. Καθηγητής Π. Λουρίδας

Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας,
Οικονομικό Πανεπιστήμιο Αθηνών
louridas@aueb.gr

Επισκόπηση

Γενικά

- Πόσο μετράει μία ψήφος;
- Στις εκλογές γενικώς ακολουθούμε την αρχή «Ένα Άτομο Μία Ψήφος» (One Person One Vote, OPOV).
- Μπορούμε εύκολα να σκεφτούμε ότι σε περίπτωση που έχουμε εκατομμύρια ψήφους, μία ψήφος μόνης της δύσκολα θα κάνει τη διαφορά.
- Δεδομένου ότι το να ψηφίσει κάποιος απαιτεί χρόνο και ότι δεν είναι σαφές πόσο μετράει μια ψήφος, το αποκαλούμενο «παράδοξο της ψηφοφορίας» αφορά ακριβώς την προθυμία μας να ψηφίσουμε ακόμα και αν δεν είναι φανερό ότι έχει σημασία.

Ευρωπαϊκή Οικονομική Κοινότητα, 1958

- Το 1958 η πρόγονη της Ευρωπαϊκής Ένωσης ήταν η Ευρωπαϊκή Οικονομική Κοινότητα (ΕΟΚ).
- Σε αυτήν συμμετείχαν οι: Γαλλία, Δυτική Γερμανία, Ιταλία, Βέλγιο, Ολλανδία, και Λουξεμβούργο.
- Ένα από τα σώματα διακυβέρνησης της ΕΟΚ ήταν το Συμβούλιο Υπουργών.
- Στο Συμβούλιο Υπουργών οι υπουργοί των χωρών ψήφιζαν για θέματα της ΕΟΚ.

Συμβούλιο Υπουργών

- Επειδή μεταξύ των χωρών υπήρχαν τεράστιες διαφορές (πληθυσμός, πλούτος, κ.λπ.) διαφορετικές χώρες είχαν διαφορετικό βάρος στους ψήφους τους.
- Η Γαλλία, Γερμανία και Ιταλία είχαν βάρος 4.
- Το Βέλγιο και η Ολλανδία είχαν βάρος 2.
- Το Λουξεμβούργο είχε βάρος 1.
- Για να περάσει μία απόφαση θα έπρεπε να συγκεντρώσει τουλάχιστον 12 ψήφους.

Πιθανές Εκβάσεις στο Συμβούλιο Υπουργών

FR (4)	DE (4)	IT (4)	NL (2)	BE (2)	LU (1)	Sum
✓	✓	✓				12
✓	✓		✓	✓		12
✓		✓	✓	✓		12
	✓	✓	✓	✓		12

Πιθανές Εκβάσεις στο Συμβούλιο Υπουργών

- Από τον πίνακα βλέπουμε ότι το όριο των 12 ψήφων επιτυγχάνεται πάντοτε και χωρίς την ψήφο του Λουξεμβούργου.
- Αυτό σημαίνει ότι το Λουξεμβούργο είχε μεν ψήφο, αλλά δεν είχε καθόλου δύναμη.
- Δεν είχε καθόλου δύναμη γιατί δεν μπορούσε να επηρεάσει την έκβαση απολύτως καμμίας ψηφοφορίας.
- Γενικώς, τι ακριβώς μας λένε τα βάρη; Από τα βάρη φαίνεται ότι η Γερμανία έχει διπλάσια ισχύ από το Βέλγιο. Ισχύει;

Εκλογικά Παίγνια

Ορισμός

- Έστω ότι έχουμε ένα σύνολο ψηφοφόρων $V = \{v_1, v_2, \ldots, v_n\}$, και ένα σύνολο βαρών $W = \{w_1, w_2, \ldots, w_m\}$.
- Ο ψηφοφόρος v_i ψηφίζει με ένα βάρος w_j, όπως ορίζεται από μια απεικόνιση $f : V \longrightarrow W$.
- Για να παρθεί μια απόφαση πρέπει να φτάσει ένα *κατώφλι* (quota) Q.
- Ο συνδυασμός των V, W, f, και Q ονομάζεται *εκλογικό παίγνιο* (voting game).

Στο παράδειγμα της ΕΟΚ έχουμε $Q = 12$.

Συνασπισμοί

Ορισμός

- Κάθε υποσύνολο ψηφοφόρων ονομάζεται *συνασπισμός* (coalition).
- Ένας συνασπισμός ο οποίος επιτυγχάνει να φτάσει το κατώφλι ονομάζεται *νικηφόρος συνασπισμός* (winning coalition).
- Ένας συνασπισμός ο οποίος δεν φτάνει το κατώφλι ονομάζεται *ηττημένος συνασπισμός* (losing coalition).

Συνασπισμοί

- Αν σε έναν νικηφόρο συνασπισμό προσθέσουμε ψηφοφόρους, παίρνουμε έναν ακόμα νικηφόρο συνασπισμό.
- Για παράδειγμα, αν στον νικηφόρο συνασπισμό {DE, FR, IT} προσθέσουμε το Βέλγιο, ο {DE, FR, IT, BE} που προκύπτει είναι πάλι νικηφόρος.
- Είναι πιο ενδιαφέρον να δούμε τι γίνεται αν από έναν νικηφόρο συνασπισμό αφαιρέσουμε έναν ψηφοφόρο.
- Για παράδειγμα, αν από τον {DE, FR, IT} αφαιρέσουμε μια χώρα, ο συνασπισμός που προκύπτει δεν είναι πλέον νικηφόρος.

Ελάχιστος Νικηφόρος Συνασπισμός

Ορισμός

Ονομάζουμε *ελάχιστο νικηφόρο συνασπισμό* (minimal winning coalition) ένα νικηφόρο συνασπισμό όταν η αφαίρεση οποιουδήποτε ψηφοφόρου από αυτόν τον κάνει να μην είναι νικηφόρος.

- Ένας ψηφοφόρος που αν αφαιρεθεί από ένα συνασπισμό αυτός παύει να είναι νικηφόρος ονομάζεται *κρίσιμος* (critical) ή swinger ή pivot.
- Σε έναν ελάχιστο νικηφόρο συνασπισμό όλοι οι ψηφοφόροι είναι κρίσιμοι.
- Ένας ψηφοφόρος μπορεί να είναι κρίσιμος σε έναν συνασπισμό που δεν είναι ελάχιστος.
- Για παράδειγμα, η DE στον {DE, FR, IT, BE}.

Ελάχιστος Νικηφόρος Συνασπισμός

- Ένας κρίσιμος ψηφοφόρος μπορεί να αλλάξει το αποτέλεσμα μιας εκλογικής αναμέτρησης.
- Αν ένας ψηφοφόρος δεν είναι κρίσιμος σε κανέναν συνασπισμό, τότε η συμπεριφορά του δεν έχει καμμία σημασία για το αποτέλεσμα των εκλογών.
- Ένας τέτοιος ψηφοφόρος ονομάζεται *μαριονέτα* (dummy).
- Στο παράδειγμα της ΕΟΚ, το Λουξεμβούργο ήταν μαριονέτα.

Βαθμολογία Banzhaf

Ορισμός

Η *βαθμολογία Banzhaf* (Banzhaf score) ενός ψηφοφόρου v_i είναι ο αριθμός των συνασπισμών στους οποίους είναι κρίσιμος. Συμβολίζουμε τη βαθμολογία Banzhaf με $\eta(v_i)$.

- Το όνομα προέρχεται από τον John F. Banzhaf III.
- Η βαθμολογία Banzhaf δεν είναι πολύ χρήσιμη γιατί δεν μας δείχνει πόσο σημαντικός είναι ο αριθμός των συνασπισμών αυτών σε σχέση με όλους τους κρίσιμους συνασπισμούς.
- Μπορεί να υπάρχουν πολλοί άλλοι κρίσιμοι συνασπισμοί στους οποίους δεν είναι κρίσιμος ο v_i.

Κατάταξη Banzhaf

Ορισμός

Κατάταξη Banzhaf (Banzhaf index of voting power, Banzhaf index) είναι ο αριθμός των συνασπισμών στους οποίους είναι κρίσιμος ο v_i προς το συνολικό αριθμό των κρίσιμων συνασπισμών για όλους τους ψηφοφόρους. Τη συμβολίζουμε με $\beta(v_i)$, οπότε έχουμε:

$$\beta(v_i) = \frac{\eta(v_i)}{\eta(v_1) + \eta(v_2) + \cdots + \eta(v_n)}$$

Αν θεωρήσουμε τη συνολική εκλογική επιρροή ως μια πίτα, η κατάταξη Banzhaf είναι το μέρος της πίτας, ο λόγος της εκλογικής επιρροής, που έχει ένας συγκεκριμένος ψηφοφόρος.

Παράδειγμα Κατάταξης Banzhaf

- Έστω τέσσερεις ψηφοφόροι $V = \{A, B, C, D\}$ με αντίστοιχα βάρη 4, 2, 1, 3 και κατώφλι $Q = 6$.
- Οι κρίσιμοι συνασπισμοί είναι οι εξής (υπογραμμίζουμε τους κρίσιμους ψηφοφόρους) $\{\underline{A}, \underline{B}\}$, $\{\underline{A}, \underline{D}\}$, $\{\underline{A}, \underline{B}, C\}$, $\{\underline{A}, B, D\}$, $\{\underline{A}, C, \underline{D}\}$, $\{\underline{B}, \underline{C}, \underline{D}\}$.
- Έχουμε λοιπόν $\eta(v_A) = 5$, $\eta(v_B) = 3$, $\eta(v_C) = 1$, και $\eta(v_D) = 3$.
- Συνεπώς βρίσκουμε ότι $\beta(v_A) = 5/12$, $\beta(v_B) = 3/12$, $\beta(v_C) = 1/12$, $\beta(v_D) = 3/12$.
- Προσέξτε ότι ο D έχει μεγαλύτερο βάρος από τον B, αλλά αυτό δεν μεταφράζεται σε μεγαλύτερη εκλογική δύναμη!
- Επίσης προσέξτε ότι με το χέρι δεν μπορούμε να υπολογίσουμε την κατάταξη Banzhaf για εκλογές με πολλούς ψηφοφόρους.

Σχετική και Απόλυτη Ισχύς Ψήφου

- Η κατάταξη Banzhaf είναι ένα *σχετικό μέγεθος* (relative measure).
- Μας δείχνει πώς σχετίζεται μια μέτρηση για κάτι σε σχέση με άλλες· για παράδειγμα, το μέρος του εισοδήματος μιας ομάδας που αντιστοιχεί σε κάθε μέλος της.
- Είναι αντίστοιχο με τη μέτρηση του ποσοστού μιας ποσότητας που αντιστοιχεί σε ένα άτομο.
- Μας ενδιαφέρει επίσης να βρούμε και ένα *απόλυτο μέγεθος* (absolute measure): πόση ποσότητα ακριβώς αντιστοιχεί σε ένα άτομο; Ποιο είναι το εισόδημα ενός ατόμου;

Συνασπισμοί και Δυναμοσύνολα

- Κάθε συνασπισμός αντιστοιχεί σε ένα υποσύνολο του συνόλου S των ψηφοφόρων.
- Το σύνολο όλων των δυνατών υποσυνόλων του S ονομάζεται *δυναμοσύνολο* (power set) και συμβολίζεται με 2^S.
- Αν το μέγεθος του S είναι n, το μέγεθος του 2^S είναι 2^n.
- Αυτό συμβαίνει γιατί κάθε υποσύνολο του S μπορούμε να το αντιστοιχίσουμε σε έναν δυαδικό αριθμό των n bits.

Δυναμοσύνολα και Δυαδικοί Αριθμοί

	x	y	z
$\varnothing$	0	0	0
$\{z\}$	0	0	1
$\{y\}$	0	1	0
$\{y, z\}$	0	1	1
$\{x\}$	1	0	0
$\{x, z\}$	1	0	1
$\{x, y\}$	1	1	0
$\{x, y, z\}$	1	1	1

Μετρική Banzhaf

- Αν κάθε συνασπισμός είναι εξίσου πιθανός, η πιθανότητα να εμφανιστεί είναι $1/2^n$.
- Αν αφαιρέσουμε έναν ψηφοφόρο v_i έχουμε $n-1$ ψηφοφόρους, άρα 2^{n-1} συνασπισμούς.
- Η πιθανότητα για έναν ψηφοφόρο να είναι κρίσιμος είναι η πιθανότητα να γίνει κρίσιμος ένας από τους 2^{n-1} συνασπισμούς με την προσθήκη του v_i.
- Αυτή η πιθανότητα είναι ίση με τον αριθμό των κρίσιμων συνασπισμών του v_i προς τον αριθμό όλων των συνασπισμών χωρίς τον v_i.

Μετρική Banzhaf

Ορισμός

Ο λόγος της βαθμολογίας Banzhaf $\eta(v_i)$ προς το 2^{n-1} ονομάζεται *μετρική Banzhaf* (Banzhaf measure) και συμβολίζεται με $\beta'(v_i)$:

$$\beta'(v_i) = \frac{\eta(v_i)}{2^{n-1}}$$

Μετρική Banzhaf

- Η μετρική Banzhaf είναι ένα απόλυτο μέγεθος.
- Η μετρική Banzhaf είναι η πιθανότητα, αν δεν ξέρουμε πώς θα ψηφίσει ο v_i, όταν μετρηθούν οι ψήφοι, αν ο v_i αλλάξει την ψήφο του, τότε να αλλάξει και το αποτέλεσμα της ψηφοφορίας.
- Επίσης είναι η πιθανότητα, αν ξέρουμε πώς θα ψηφίσει ο v_i, να αλλάξει το αποτέλεσμα των εκλογών αν αλλάξει γνώμη ο v_i.

Παράδειγμα Μετρικής Banzhaf

Συνασπισμοί χωρίς A	Συνασπισμοί με A	Ψήφοι	Νίκη	Κρίσιμος
$\varnothing$	$\{A\}$	4		
$\{B\}$	$\{A, B\}$	6	✓	✓
$\{C\}$	$\{A, C\}$	5		
$\{D\}$	$\{A, D\}$	7	✓	✓
$\{B, C\}$	$\{A, B, C\}$	7	✓	✓
$\{B, D\}$	$\{A, B, D\}$	9	✓	✓
$\{C, D\}$	$\{A, C, D\}$	8	✓	✓
$\{B, C, D\}$	$\{A, B, C, D\}$	10	✓	

Παράδειγμα Μετρικής Banzhaf

- Η μετρική Banzhaf για τον A είναι $5/8$.
- Για τους υπόλοιπους έχουμε $B = 3/8$, $C = 1/8$, $D = 3/8$.
- Αφού η μετρική δεν είναι κανονικοποιημένη, δεν αθροίζει στο ένα.
- Αυτό συμβαίνει επειδή η μετρική Banzhaf είναι απόλυτο μέγεθος, όπως θέλαμε.
- Με αυτήν μπορούμε να ελέγξουμε την ισχύ ενός ψηφοφόρου σε διαφορετικές εκλογές, πράγμα που δεν μπορούμε να το κάνουμε με την κατάταξη Banzhaf.

Μετρική Banzhaf και Κατάταξη Banzhaf

- Μπορούμε από τη μετρική Banzhaf να πάμε στην κατάταξη Banzhaf ως εξής:

$$\beta(v_i) = \beta'(v_i) \times \frac{2^{n-1}}{\eta(v_1) + \eta(v_2) + \cdots + \eta(v_n)}$$

- Συνεπώς μπορούμε να θεωρήσουμε τη μετρική Banzhaf ως το βασικό μας μέγεθος και από εκεί να εξάγουμε την κατάταξη Banzhaf αν θέλουμε.

Υπολογισμός Μετρικής Banzhaf

- Η διαδικασία της απαρίθμησης που χρησιμοποιήσαμε για να υπολογίσουμε το $\beta'(v_i) = \eta(v_i)/2^{n-1}$ μπορεί να λειτουργήσει μόνο όσο ο αριθμός των ψηφοφόρων είναι μικρός.
- Ο υπολογισμός του $\beta'(v_i)$ για μεγάλο αριθμό ψηφοφόρων είναι δύσκολος.
- Ο παρονομαστής δεν είναι πρόβλημα—παρά το ότι το 2^{n-1} είναι ένας πολύ μεγάλος αριθμός, μπορούμε να τον υπολογίσουμε ακριβώς.
- Το πρόβλημα είναι ο αριθμητής. Δεν υπάρχει τρόπος να υπολογίσουμε το $\eta(v_i)$ αποτελεσματικά.

Πιθανοθεωρητικός Υπολογισμός Μετρικής Banzhaf

- Μπορούμε όμως να υπολογίσουμε το $\beta'(v_i)$ πιθανοθεωρητικά.
- Έστω ότι παίρνουμε συνασπισμούς στην τύχη.
- Κάποιοι από αυτούς θα είναι κρίσιμοι.
- Αν τους παίρνουμε πραγματικά στην τύχη, και έχουμε ένα επαρκές δείγμα, τότε ο λόγος των κρίσιμων συνασπισμών προς όλους τους συνασπισμούς του δείγματός μας θα προσεγγίζει τον πραγματικό λόγο.

Επισκόπηση

Μέθοδος Monte Carlo

Ορισμός

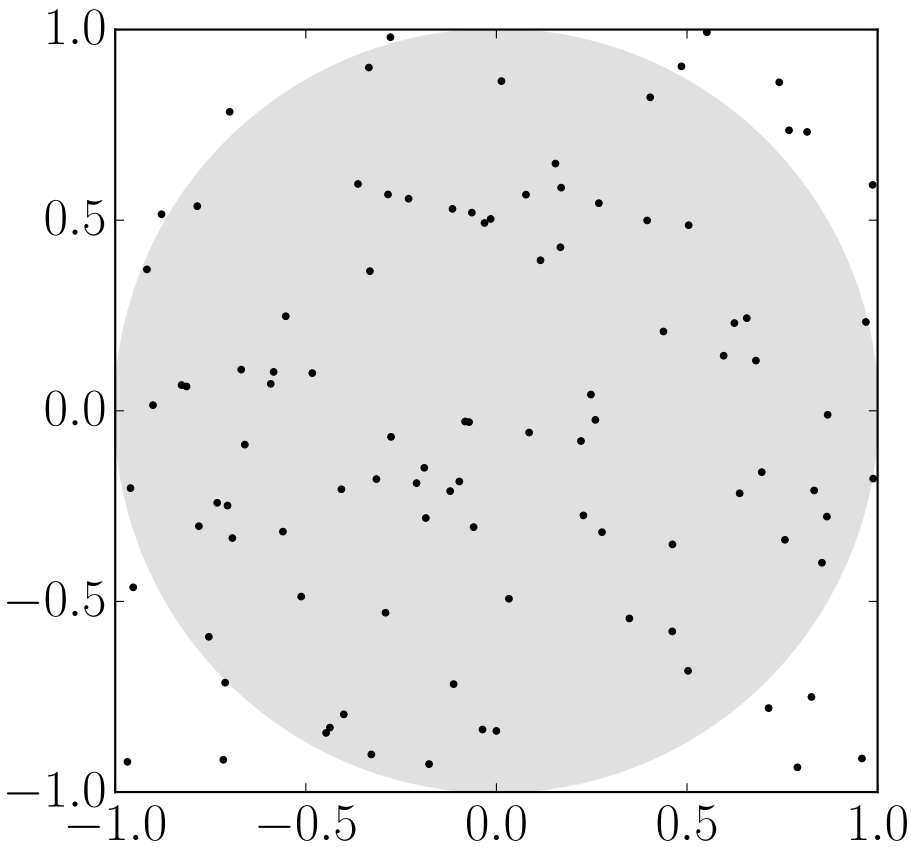
Μια υπολογιστική μέθοδος η οποία χρησιμοποιεί τυχαία δειγματοληψία για την παραγωγή των αποτελεσμάτων της ονομάζεται *μέθοδος Monte Carlo*.

- Ονομάστηκε από το διάσημο καζίνο.
- Εφευρέθηκε από τους πρωτοπόρους των ψηφιακών υπολογιστών.
- Χρησιμοποιείται στις φυσικές επιστήμες, από μηχανικούς, μέχρι και τα χρηματοοικονομικά.

Μέθοδος Monte Carlo: Υπολογισμός π

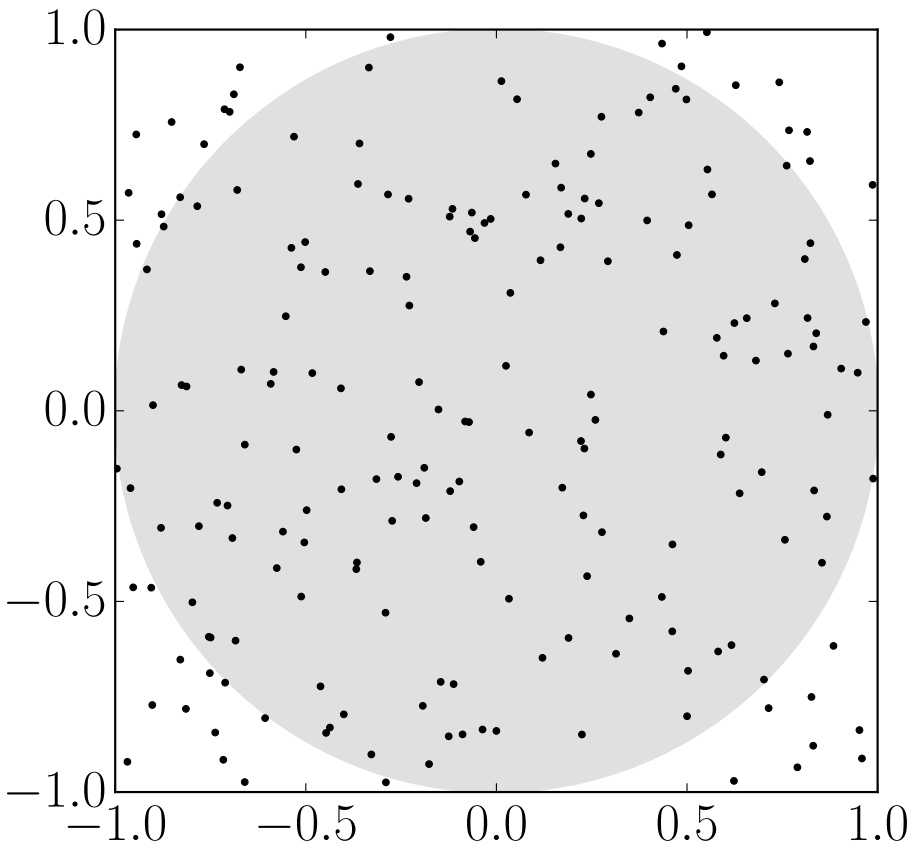
- Έστω ότι έχουμε ένα τετράγωνο με μήκος πλευράς 2 μονάδες, το εμβαδόν του θα είναι 4 τετραγωνικές μονάδες.
- Αν εγγράψουμε έναν κύκλο μέσα στο τετράγωνο, η διάμετρός του θα είναι 2 μονάδες και η ακτίνα του θα είναι 1 μονάδα.
- Συνεπώς το εμβαδόν του κύκλου θα είναι π.
- Αν ρίξουμε τυχαία αντικείμενα στο τετράγωνο, π.χ. κόκκους ριζιού, κάποια θα πέσουν μέσα στον κύκλο, κάποια έξω.
- Αν ρίξουμε αρκετά, ο λόγος αυτών μέσα στον κύκλο προς το λόγο όλων θα είναι $\pi/4$.

Μέθοδος Monte Carlo: Υπολογισμός π



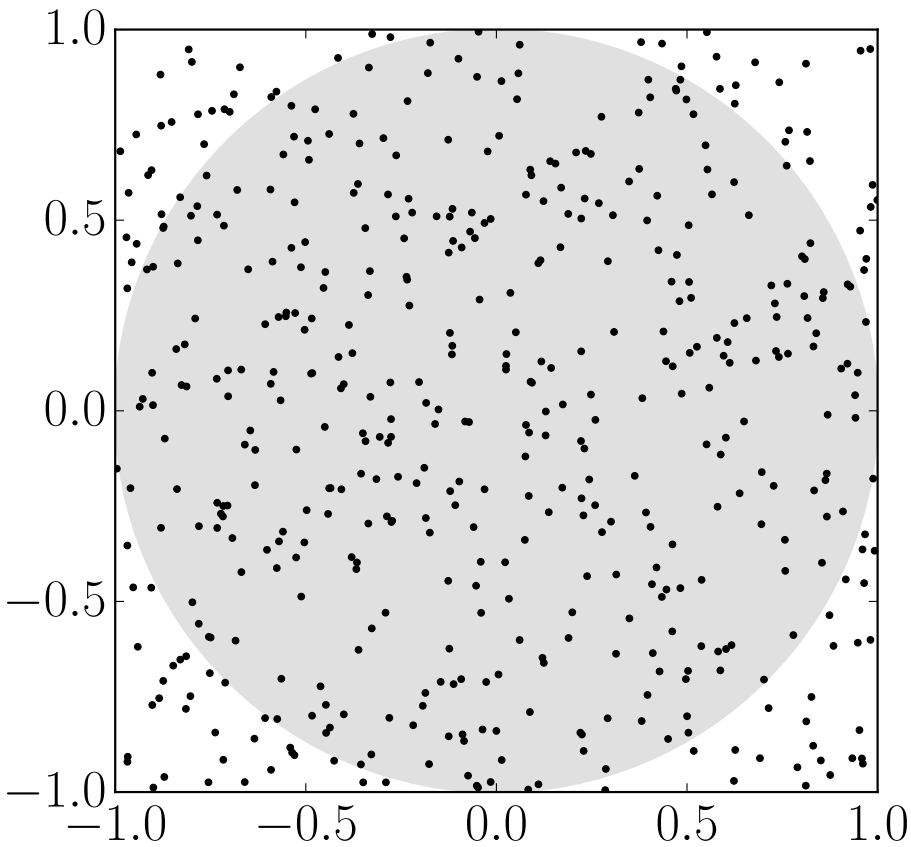
100 σημεία, $\pi = 3{,}2$, $s_e = 0{,}017$.

Μέθοδος Monte Carlo: Υπολογισμός π



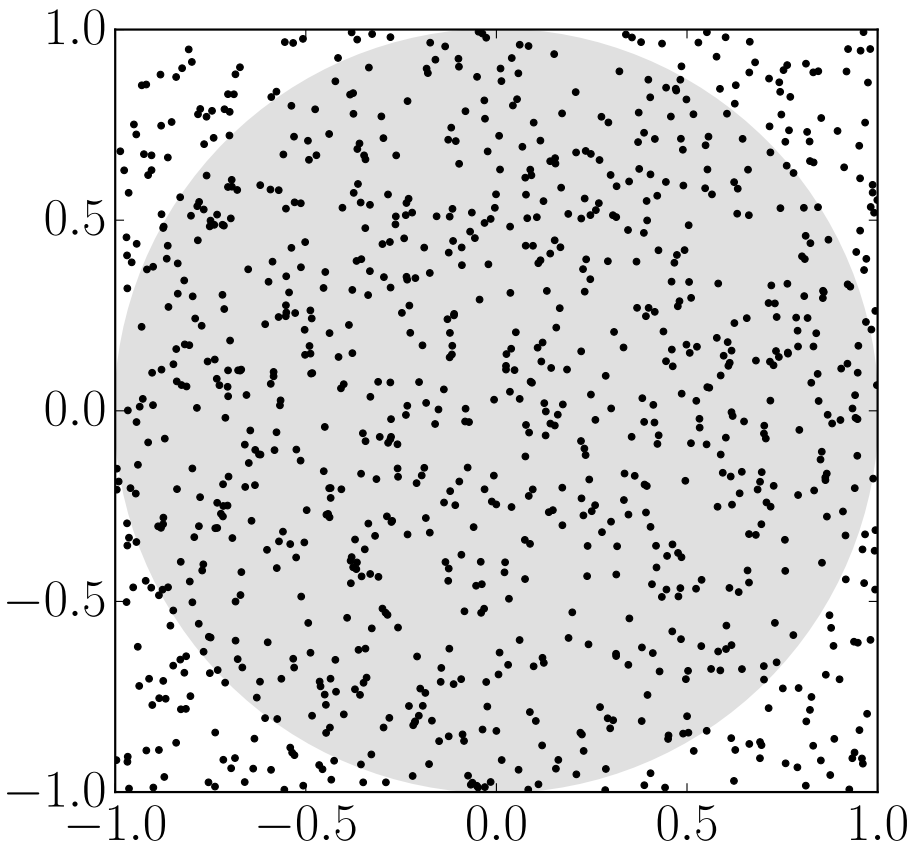
200 σημεία, $\pi = 3{,}12$, $s_e = 0{,}012$.

Μέθοδος Monte Carlo: Υπολογισμός π



500 σημεία, $\pi = 3{,}12$, $s_e = 0{,}008$.

Μέθοδος Monte Carlo: Υπολογισμός π



1000 σημεία, $\pi = 3{,}14$, $s_e = 0{,}005$.

Μέθοδος Monte Carlo: Υπολογισμός π—Τυπικό Σφάλμα

- Ορίζουμε μια μεταβλητή X ίση με 1 όταν το σημείο είναι μέσα στον κύκλο και ίση με 0 όταν είναι εκτός.
- Η *αναμενόμενη τιμή* (expected value) της X είναι $E[X] = \pi/4 \times 1 + (1 - \pi/4) \times 0 = \pi/4$.
- Η *διακύμανση* (variance) της X είναι $\sigma^2 = E[X^2] - (E[X])^2$.
- Έχουμε $E[X^2] = (\pi/4) \times 1^2 + [1 - (\pi/4)] \times 0^2 = \pi/4$. Άρα $\sigma^2 = \pi/4 - (\pi/4)^2 = (\pi/4)(1 - \pi/4)$.
- Το *τυπικό σφάλμα* (standard error) τότε είναι $s_e = \sigma/\sqrt{n}$.
- Το σ είναι η *τυπική απόκλιση* (standard deviation).
- Ένα τυπικό σφάλμα s_e σημαίνει ότι έχουμε πιθανότητα 95% η πραγματική τιμή του π να βρίσκεται μέσα σε ένα διάστημα $\pm 1{,}96 s_e$ από την τιμή που υπολογίσαμε.

Μέθοδος Monte Carlo: Υπολογισμός π—Μόνο για Παράδειγμα!

Προσοχή, η μέθοδος που περιγράψαμε είναι βολική για να δούμε πώς δουλεύει η μέθοδος Monte Carlo, αλλά στην πραγματικότητα υπάρχουν πολύ πιο αποτελεσματικές μέθοδοι για τον υπολογισμό του π.

Επισκόπηση

Monte Carlo Υπολογισμός Μετρικής Banzhaf

- Μπορούμε να εφαρμόσουμε τη μέθοδο Monte Carlo για τον υπολογισμό της μετρικής Banzhaf.
- Θα δημιουργούμε τυχαίους συνασπισμούς και θα ελέγχουμε αν είναι κρίσιμοι.
- Ο λόγος αυτών που είναι κρίσιμοι προς το σύνολο αυτών που παράγουμε θα είναι η μετρική Banzhaf.
- Για να το κάνουμε αυτό χρειαζόμαστε έναν αλγόριθμο για την παραγωγή τυχαίων υποσυνόλων ενός συνόλου.

Αλγόριθμος Παραγωγής Τυχαίου Υποσυνόλου

Algorithm: Random subset generation.

`RandomSubset(`S`)` $\longrightarrow$ RS

Input: S, a set

Output: RS, a random subset of S

```
1  RS ← CreateList()
2  foreach m in S do
3      r ← Random(0, 1)
4      if r < 0.5 then
5          InsertInList(RS, NULL, m)
6  return RS
```

Αλγόριθμος Παραγωγής Τυχαίου Υποσυνόλου

- Η είσοδος είναι ένα σύνολο S.
- Η έξοδος RS είναι ένα τυχαίο υποσύνολο του S.
- Εξετάζουμε κάθε στοιχείο του S και αποφασίζουμε τυχαία αν θα το συμπεριλάβουμε στο RS ή όχι.
- Για να το αποφασίσουμε παίρνουμε έναν τυχαίο αριθμό στο $[0, 1)$ και τον συγκρίνουμε με το 0,5. Αν είναι μικρότερος του 0,5 τον συμπεριλαμβάνουμε στο RS, αλλιώς το αφήνουμε.

Monte Carlo Υπολογισμός Μετρικής Banzhaf

Algorithm: Monte Carlo Banzhaf measure.

`BanzhafMeasure`$(v, ov, q, w, t) \rightarrow b$

Input: v, a voter
ov, a list containing the other voters
q, the quota required
w an associative array containing the weight of each voter
t, the number of tries

Output: b, the Banzhaf measure for voter v

```
1  k ← 0
2  nc ← 0
3  while k < t do
4      coalition ← RandomSubset(ov)
5      votes ← 0
6      foreach m in coalition do
7          votes ← votes + Lookup(w, m)
8      if votes < q and votes + Lookup(w, v) ≥ q then
9          nc ← nc + 1
10     k ← k + 1
11 b ← nc/k
12 return b
```

Monte Carlo Υπολογισμός Μετρικής Banzhaf

- Ο αλγόριθμος υπολογίζει τη μετρική Banzhaf για ένα ψηφοφόρο εκτελώντας t επαναλήψεις.
- Για κάθε επανάληψη, παίρνουμε ένα τυχαίο υποσύνολο, δηλαδή συνασπισμό, που δεν περιέχει τον ψηφοφόρο για τον οποίο υπολογίζουμε τη μετρική.
- Υπολογίζουμε το άθροισμα των ψήφων του υποσυνόλου.
- Αν προσθέτοντας και τους ψήφους του ψηφοφόρου περνάμε το κατώφλι, ο συνασπισμός είναι κρίσιμος.
- Στο τέλος διαιρούμε τους κρίσιμους συνασπισμούς προς το σύνολο των συνασπισμών.

Ακρίβεια Monte Carlo Υπολογισμού Μετρικής Banzhaf

- Θέλουμε να ξέρουμε τον αριθμό των επαναλήψεων που απαιτούνται για να πετύχουμε μια συγκεκριμένη ακρίβεια.
- Προκύπτει ότι αν θέλουμε το αποτέλεσμά μας να βρίσκεται μέσα σε ένα διάστημα $\pm\epsilon$ με πιθανότητα δ, ο απαιτούμενος αριθμός δειγμάτων είναι:

$$k \geq \frac{\ln \frac{2}{1-\delta}}{2\epsilon^2}$$

Μετρική Banzhaf Αμερικάνικων Πολιτειών

- Με βάση τα παραπάνω μπορούμε να υπολογίσουμε τη μετρική Banzhaf των αμερικάνικων πολιτειών.
- Στις ΗΠΑ ο Πρόεδρος εκλέγεται από το κολλέγιο των εκλεκτόρων. Οι εκλέκτορες προέρχονται από κάθε πολιτεία συν την περιοχή της πρωτεύουσας (District of Columbia).
- Συνολικά υπάρχουν 538 εκλέκτορες, άρα το κατώφλι είναι 270 εκλέκτορες. Ο αριθμός εκλεκτόρων ανά πολιτεία προκύπτει από την απογραφή του πληθυσμού.
- Τα αποτελέσματα που ακολουθούν υπολογίστηκαν με $\epsilon = 0{,}001$ και $\delta = 0{,}95$.
- Απαιτήθηκαν 1.844.440 δείγματα για κάθε μετρική.

Μετρική Banzhaf Αμερικάνικων Πολιτειών

CA	55	0.471	MN	10	0.076	NM	5	0.038
TX	38	0.298	MO	10	0.075	WV	5	0.038
FL	29	0.223	WI	10	0.076	HI	4	0.03
NY	29	0.224	AL	9	0.068	ID	4	0.03
IL	20	0.153	CO	9	0.068	ME	4	0.03
PA	20	0.153	SC	9	0.068	NH	4	0.03
OH	18	0.136	KY	8	0.06	RI	4	0.03
GA	16	0.121	LA	8	0.061	AK	3	0.023
MI	16	0.121	CT	7	0.053	DC	3	0.023
NC	15	0.114	OK	7	0.052	DE	3	0.023
NJ	14	0.106	OR	7	0.053	MT	3	0.023
VA	13	0.098	AR	6	0.045	ND	3	0.023
WA	12	0.091	IA	6	0.045	SD	3	0.023
AZ	11	0.083	KS	6	0.045	VT	3	0.023
IN	11	0.083	MS	6	0.045	WY	3	0.023
MA	11	0.083	NV	6	0.045			
TN	11	0.083	UT	6	0.046			
MD	10	0.076	NE	5	0.038			

Επισκόπηση

Γενικά

- Σε πολλές κρυπτογραφικές εφαρμογές χρειάζεται να βρούμε μεγάλους πρώτους αριθμούς.
- Οι αριθμοί αυτοί έχουν συνήθως m bits, όπου το m είναι μια μεγάλη δύναμη του 2 (1024, 2048, 4096, ...).
- Η εύρεση πρώτων αριθμών δεν είναι εύκολη υπόθεση.
- Σύμφωνα με το Θεώρημα των Πρώτων Αριθμών (Prime Number Theorem), αν n ένας μεγάλος αριθμός, ο αριθμός των πρώτων αριθμών μικρότερων ή ίσων με το n είναι περίπου $n/\ln n$.
- Αλλά πώς τους βρίσκουμε;

Το Κόσκινο του Ερατοσθένη

- Ένας τρόπος είναι να βρούμε όλους τους πρώτους αριθμούς με m bits και να επιλέξουμε έναν από αυτούς.
- Αυτό μπορούμε να το κάνουμε με το *κόσκινο του Ερατοσθένη*.
- Ξεκινάμε με τον 2, που είναι πρώτος.
- Σημειώνουμε όλα τα πολλαπλάσια του 2 μέχρι το όριο που έχουμε θέσει—αυτά φυσικά δεν είναι πρώτοι.
- Επιστρέφουμε στο 2 και αναζητούμε τον πρώτο αριθμό μεγαλύτερο του 2 που δεν τον έχουμε σημειώσει ως σύνθετο.
- Σημειώνουμε πάλι όλα τα πολλαπλάσια αυτού του αριθμού γιατί είναι σύνθετοι.
- Συνεχίζουμε έτσι μέχρι να εξαντλήσουμε όλους τους δυνατούς αριθμούς.

Το Κόσκινο του Ερατοσθένη

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	F	F	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
2	F	F	T	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T
3	F	F	T	T	F	T	F	T	F	F	F	T	F	T	F	F	F	T	F	T	F	F	F	T	F	T	F	F	F	T	F	T
5	F	F	T	T	F	T	F	T	F	F	F	T	F	T	F	F	F	T	F	T	F	F	F	T	F	F	F	F	F	T	F	T

Ο Αλγόριθμος του Κόσκινου του Ερατοσθένη

Algorithm: The Sieve of Eratosthenes.

`SieveEratosthenes`$(n) \rightarrow$ *isprime*

Input: n, a natural number greater than 1

Output: *isprime*, a boolean array of size $n+1$ such that if $p \leq n$ is a prime, *isprime*$[p]$ is TRUE, otherwise it is FALSE

```
1  isprime ← CreateArray(n + 1)
2  isprime[0] ← FALSE
3  isprime[1] ← FALSE
4  for i ← 2 to n + 1 do
5      isprime[i] ← TRUE
6  p ← 2
7  while p^2 ≤ n do
8      if isprime[p] = TRUE then
9          j ← p
10         while j ≤ ⌊n/p⌋ do
11             isprime[j × p] ← FALSE
12             j ← j + 1
13     p ← p + 1
14 return isprime
```

Ανάλυση Αλγορίθμου

- Σε κάθε επανάληψη του εξωτερικού βρόχου σημειώνουμε ως σύνθετους όλα τα πολλαπλάσια των πρώτων αριθμών $p \leq \sqrt{n}$.
- Η τετραγωνική ρίζα εμφανίζεται γιατί κάθε σύνθετος c τέτοιος ώστε $\sqrt{n} \leq c \leq n$ μπορεί να γραφτεί ως το γινόμενο δύο παραγόντων $c = f_1 \times f_2$ όπου ισχύει τουλάχιστον ένα των:
  - $f_1 \leq \sqrt{n}$
  - $f_2 \leq \sqrt{n}$

  ($f_1 = f_2 = \sqrt{n}$ αν το n είναι τέλειο τετράγωνο).
- Στην πρώτη επανάληψη σημειώνουμε τα πολλαπλάσια του 2, $\lfloor n/2 \rfloor \cdot$ στη δεύτερη επανάληψη τα πολλαπλάσια του 3, $\lfloor n/3 \rfloor \cdot$ μετά όλα τα πολλαπλάσια του 5, $\lfloor n/5 \rfloor$, μέχρι τα πολλαπλάσια του μεγαλύτερου πρώτου $k < \sqrt{n}$.

Ανάλυση Αλγορίθμου (Συνέχεια)

- Συνολικά σημειώνουμε $n/2 + n/3 + n/5 + \cdots + n/k$ σύνθετους, που είναι ίσο με $n(1/2 + 1/3 + 1/5 + \cdots + 1/k)$.
- Το άθροισμα $(1/2 + 1/3 + 1/5 + \cdots + 1/k)$ είναι το άθροισμα των αντιστρόφων των πρώτων που δεν είναι μεγαλύτεροι από το $\sqrt{n}$.
- Αποδεικνύεται ότι το άθροισμα των αντιστρόφων των πρώτων που δεν είναι μεγαλύτεροι από έναν αριθμό m είναι $O(\log \log m)$.
- Άρα ο συνολικός χρόνος που καταναλώνουμε σημειώνοντας σύνθετους είναι $O(n \log \log \sqrt{n}) = O(n \log \log n)$, που μας δίνει και την πολυπλοκότητα του αλγορίθμου.

Αποτελεσματικότητα Εύρεσης Πρώτων

- Υπάρχουν ακόμα καλύτεροι αλγόριθμοι οι οποίοι έχουν πολυπλοκότητα $O(n)$.
- Φαινομενικά λοιπόν δεν έχουμε πρόβλημα με την πολυπλοκότητα.
- Δεν είναι έτσι όμως.
- Είπαμε ότι $n = 2^m$, συνεπώς στην πραγματικότητα η πολυπλοκότητα είναι $O(2^m)$.
- Για έναν αριθμό των 4096 bits παίρνουμε $O(2^{4096})$, που δεν είναι πρακτικό.

Εύρεση Πρώτων Πιθανοθεωρητικά

- Από το Θεώρημα Πρώτων Αριθμών, αν εξετάσουμε και τους n αριθμούς περιμένουμε να βρούμε $n/\ln n$ πρώτους.
- Αν πάρουμε έναν αριθμό στην τύχη, η πιθανότητα να είναι πρώτος είναι $1/\ln n$.
- Αντίστροφα, προκύπτει ότι προκειμένου να βρούμε έναν πρώτο θα πρέπει να δοκιμάσουμε κατά μέσο όρο $\ln n$ αριθμούς.
- Αν ψάχνουμε πρώτους αριθμούς με 4096 bits, θα πρέπει να δοκιμάσουμε κατά μέσο όρο $\ln(2^{4096}) \approx 2840$ αριθμούς.
- Αν έχουμε έναν τρόπο να ελέγχουμε γρήγορα αν ένας αριθμός είναι πρώτος, η προσέγγιση αυτή είναι εφικτή.

Έλεγχος Πρώτου Αριθμού

- Για να ελέγξουμε αν ένας αριθμός n είναι πρώτος μπορούμε να τον διαιρέσουμε με όλους τους αριθμούς από το 2 μέχρι το $\sqrt{n}$.
- Αυτό γιατί ένας αριθμός μεγαλύτερος του $\lceil \sqrt{n} \rceil$ μπορεί να μας δώσει το n μόνο αν πολλαπλασιαστεί με έναν αριθμό που δεν ξεπερνά το $\lfloor \sqrt{n} \rfloor$.
- Στην πραγματικότητα μπορούμε να μειώσουμε στο μισό τις δοκιμές, δοκιμάζοντας μόνο περιττούς αριθμούς.
- Πάντως σε κάθε περίπτωση η προσέγγιση θέλει $O\left((1/2)\sqrt{n}\right) = O(\sqrt{n})$ βήματα.
- Αυτό είναι πάλι απαγορευτικό αφού $O\left(\sqrt{2^m}\right) = O\left((2^m)^{1/2}\right) = O(2^{m/2})$.

Πιθανοθεωρητικός Έλεγχος Πρώτου Αριθμού

- Αντί γι' αυτό θα ακολουθήσουμε μία πιθανοθεωρητική προσέγγιση.
- Θα δούμε έναν αλγόριθμο ο οποίος θα μπορεί να μας δώσει γρήγορα και με μεγάλη βεβαιότητα έναν πρώτο αριθμό.
- Ο αλγόριθμος ενδέχεται όμως να κάνει λάθος!
- Δηλαδή, να μας δώσει ένα σύνθετο αριθμό, λέγοντάς μας ότι είναι πρώτος.
- Αυτό μπορούμε να το ελέγξουμε όμως, αν η πιθανότητα να κάνει ένα τέτοιο λάθος είναι απειροελάχιστη.

Πιθανοθεωρητικός Έλεγχος Πρώτου Αριθμού

- Για να το κάνουμε αυτό, θα χρειαστούμε κατ' αρχήν έναν βοηθητικό αλγόριθμο, που θα τον ονομάσουμε *μάρτυρα* (witness).
- Αυτός θα παίρνει σαν είσοδο έναν αριθμό.
- Αν μας πει ότι ο αριθμός είναι σύνθετος, λέει σίγουρα την αλήθεια.
- Αν μας πει ότι ο αριθμός είναι πρώτος, μπορεί να κάνει λάθος.

Μάρτυρας Σύνθετου Αριθμού

- Ο μάρτυρας δεν λέει πάντα την αλήθεια.
- Αν πει: «Ο αριθμός είναι σύνθετος» μπορούμε να τον εμπιστευτούμε.
- Αν πει: «Ο αριθμός είναι πρώτος» δεν μπορούμε να είμαστε σίγουροι—αλλά θα δούμε ότι το πιθανότερο είναι ο μάρτυρας να λέει την αλήθεια.

Λειτουργία Μάρτυρα

1. Ξεκινάμε με έναν περιττό αριθμό p για τον οποίο θέλουμε να αποφανθεί ο μάρτυρας.
2. Ο αριθμός αυτός μπορεί πάντα να γραφτεί στη μορφή $p = 1 + 2^r q$ όπου q ένας περιττός αριθμός.
3. Στη συνέχεια, παίρνουμε έναν άλλο τυχαίο αριθμό x στο διάστημα $[2, p-1]$.
4. Υπολογίζουμε τον $y = x^q \bmod p$
5. Αν $y = 1$ ή $y = p - 1$ τερματίζουμε λέγοντας ότι ο p είναι πιθανώς πρώτος.
6. Διαφορετικά, υπολογίζουμε το $y \leftarrow y^2 \bmod p$.
7. Αν $y = p - 1$ τερματίζουμε λέγοντας ότι ο p είναι πιθανώς πρώτος.
8. Αν $y = 1$ τερματίζουμε λέγοντας ότι ο p είναι σίγουρα σύνθετος.
9. Διαφορετικά, επιστρέφουμε στο βήμα 6, κάνοντας μέχρι r επαναλήψεις.
10. Αν έχουμε κάνει r επαναλήψεις, τερματίζουμε λέγοντας ότι ο p είναι σίγουρα σύνθετος.

Μάρτυρας Σύνθετου Αριθμού

Algorithm: A witness for composite numbers.

`WitnessComposite`$(p) \longrightarrow$ TRUE or FALSE

Input: p, an odd integer

Output: a boolean value that is TRUE if the number is definitely composite, FALSE otherwise

```
1  (r, q) ← FactorTwo(p − 1)
2  x ← RandomInt(2, p − 1)
3  y ← x^q mod p
4  if y = 1 then
5      return FALSE
6  for j ← 0 to r do
7      if y = p − 1 then
8          return FALSE
9      y ← y^2 mod p
10     if y = 1 then
11         return TRUE
12 return TRUE
```

Αξιοπιστία Μάρτυρα Σύνθετου Αριθμού

- Ποια είναι η πιθανότητα να κάνει λάθος ο μάρτυρας;
- Αποδεικνύεται ότι η πιθανότητα ο μάρτυρας να μας πει ότι ένας αριθμός είναι πρώτος χωρίς να είναι δεν ξεπερνάει το $1/4$.
- Άρα αν τον χρησιμοποιήσουμε δύο φορές, η πιθανότητα να μας πει ότι ο p είναι πρώτος χωρίς να είναι είναι $(1/4)^2$.
- Αν τον χρησιμοποιήσουμε t φορές, η πιθανότητα να μας πει ότι ο p είναι πρώτος χωρίς να είναι είναι $(1/4)^t$.
- Για $t = 50$ η πιθανότητα να κάνει λάθος είναι $(1/4)^{50}$, που γενικώς αρκεί για πρακτικές εφαρμογές.

Έλεγχος Πρώτων Αριθμών Miller-Rabin

- Με βάση όσα είδαμε, φτάνουμε στον αλγόριθμο ελέγχου πρώτων αριθμών Miller-Rabin.
- Ο αλγόριθμος αυτός παίρνει έναν τυχαίο περιττό αριθμό ως είσοδο.
- Ελέγχει, με τη χρήση του μάρτυρα, αν ο αριθμός είναι σύνθετος.
- Αν είναι, επιστρέφει αμέσως λέγοντας ότι δεν είναι πρώτος.
- Διαφορετικά, ο μάρτυρας μπορεί να πλανάται με πιθανότητα μέχρι $(1/4)$, άρα ο αλγόριθμος καλεί εκ νέου το μάρτυρα. Η διαδικασία επαναλαμβάνεται όσες φορές θέλουμε, έστω t.
- Για t επαναλήψεις η πιθανότητα να κάνει ο αλγόριθμος λάθος είναι $(1/4)^t$.
- Η συνολική πολυπλοκότητα του μάρτυρα είναι $O((\lg p)^3)$.
- Για t επαναλήψεις θα έχουμε $O(t \cdot (\lg p)^3)$.
- Περιμένουμε να ελέγξουμε περίπου $\ln p$ αριθμούς μέχρι να βρούμε έναν πρώτο, όπου κάθε έλεγχος απαιτεί $O(t \cdot (\lg p)^3)$.

Έλεγχος Πρώτων Αριθμών Miller-Rabin

Algorithm: Miller-Rabin primality test.

`MillerRabinPrimalityTest`$(p, t) \longrightarrow$ TRUE or FALSE

Input: p, an odd integer
t, the number of times the witness primality function will be applied

Output: TRUE if the number is prime with error probability at most $(1/4)^t$, FALSE if the number is definitely composite

```
1  for i ← 0 to t do
2      if WitnessComposite(p) then
3          return FALSE
4  return TRUE
```

Εύρεση του $p = 1 + 2^r q$

- Στον αλγόριθμο του μάρτυρα αναφέραμε ότι ένας περιττός αριθμός p μπορεί να γραφτεί στη μορφή $p = 1 + 2^r q$ με q περιττό.
- Για το σκοπό αυτό χρησιμοποιήσαμε τη συνάρτηση `FactorTwo`.
- Ώρα να δούμε πώς υλοποιείται.
- Αν $p = 1 + 2^r q$, ο $2^r q$ είναι άρτιος.
- Ο παρακάτω αλγόριθμος υλοποιεί τον τρόπο εύρεσης αυτής της αναπαράστασης ενός άρτιου αριθμού n ως $2^r q$ με q περιττό.

Ανάλυση Αριθμού σε μορφή $2^r q$ με q Περιττό

Algorithm: Factor n as $2^r q$, with q odd.

`FactorTwo`$(n) \to (r, q)$

Input: n, an even integer

Output: (r, q), such that $n = 2^r q$ and q is odd

```
1  q ← n
2  r ← 0
3  while q mod 2 = 0 do
4      r ← r + 1
5      q ← q/2
6  return (r, q)
```