Θεωρία Πιθανοτήτων, εαρινό εξάμηνο 2017-18.

Λύσεις του τρίτου φυλλαδίου ασκήσεων.

1. *Θεωρούμε n κάλπες και υποθέτουμε ότι η k κάλπη περιέχει k άσπρες και $n-k$ μαύρες μπάλες. Επιλέγουμε στην τύχη μια κάλπη και τραβάμε διαδοχικά και χωρίς επανάθεση δυο μπάλες. Ποιά είναι πιθανότητα να είναι και οι δυο άσπρες;*
Συμβολίζουμε K_k το ενδεχόμενο "επιλέγουμε την k κάλπη από τις n". Τότε $P(K_k) = \frac{1}{n}$.
Ας θεωρήσουμε ότι στο πρώτο βήμα έχουμε επιλέξει την k κάλπη, και ας δούμε ποιά είναι η πιθανότητα να επιλέξουμε δύο άσπρες μπάλες στα δύο πρώτα τραβήγματα από την συγκεκριμένη κάλπη. Συμβολίζουμε A_1 το ενδεχόμενο "στο πρώτο τράβηγμα βρίσκουμε άσπρη μπάλα" και A_2 το ενδεχόμενο "στο δεύτερο τράβηγμα βρίσκουμε άσπρη μπάλα" (από την k κάλπη, εννοείται). Μας ενδιαφέρει η πιθανότητα $P(A_1 \cap A_2)$ και την βρίσκουμε ως εξής:
$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1) = \frac{k}{n}\frac{k-1}{n-1}.$$
Πράγματι, έχουμε πιθανότητα $P(A_1) = \frac{k}{n}$ να τραβήξουμε άσπρη μπάλα, όταν στην k κάλπη υπάρχουν n μπάλες από τις οποίες k είναι άσπρες. Και, με δεδομένο ότι στο πρώτο τράβηγμα βρήκαμε άσπρη μπάλα, έχουμε πιθανότητα $P(A_2|A_1) = \frac{k-1}{n-1}$ να τραβήξουμε άσπρη μπάλα όταν στην κάλπη έχουν απομείνει $n-1$ μπάλες από τις οποίες $k-1$ είναι άσπρες.
Τώρα συμβολίζουμε AA το ενδεχόμενο "τραβάμε δύο άσπρες μπάλες" και θέλουμε να βρούμε την πιθανότητα $P(AA)$. Η πιθανότητα που έχουμε υπολογίσει μέχρι στιγμής είναι η πιθανότητα να προκύψει το AA με δεδομένο ότι έχουμε επιλέξει την k κάλπη. Δηλαδή:
$$P(AA|K_k) = \frac{k}{n}\frac{k-1}{n-1}.$$
Επίσης, έχουμε και την πιθανότητα $P(K_k) = \frac{1}{n}$. Από τον τύπο ολικής πιθανότητας έχουμε
$$\begin{aligned} P(AA) &= P(AA|K_1)P(K_1) + P(AA|K_2)P(K_2) + \cdots + P(AA|K_n)P(K_n) \\ &= \sum_{k=1}^{n} P(AA|K_k)P(A_k) = \sum_{k=1}^{n}\frac{k}{n}\frac{k-1}{n-1}\frac{1}{n} = \frac{1}{n^2(n-1)}\sum_{k=1}^{n} k(k-1). \end{aligned}$$

2. *Η λέξη "και" αποτελεί το 10% των λέξεων των κειμένων της συγγραφέως Νταμά και το 20% των λέξεων των κειμένων της συγγραφέως Μηδουλίδου. Ένα κείμενο που βρέθηκε εν έτει 2532 μ.Χ. αποδίδεται κατά 60% στην Νταμά και κατά 40% στην Μηδουλίδου. Αν σε ένα τυχαίο δείγμα εννέα λέξεων του κειμένου βρέθηκε ένα "και", ποιά είναι η πιθανότητα το κείμενο να είναι της Νταμά;*
Γνωρίζουμε ότι το ενδεχόμενο N, δηλαδή "το κείμενο είναι της Νταμά", έχει πιθανότητα $P(N) = \frac{60}{100} = \frac{6}{10}$. Επίσης, ότι το ενδεχόμενο M, δηλαδή "το κείμενο είναι της Μηδουλίδου", έχει πιθανότητα $P(M) = \frac{40}{100} = \frac{4}{10}$.
Επίσης, μπορούμε να βρούμε την πιθανότητα του ενδεχομένου D, δηλαδή "το τυχαίο δείγμα εννέα λέξεων του κειμένου περιέχει ένα "και" ", με δεδομένο το ενδεχόμενο N. Πώς; Όταν επιλέγουμε τυχαία μία λέξη από το οποιοδήποτε κείμενο της Νταμά, η πιθανότητα να προκύψει η λέξη "και" είναι $\frac{10}{100} = \frac{1}{10}$ και η πιθανότητα να μην προκύψει η λέξη "και" είναι $\frac{90}{100} = \frac{9}{10}$. Άρα η πιθανότητα να προκύψει δείγμα
$$K - - - - - - - -$$
με "και" στην πρώτη θέση και "όχι και" στις άλλες οκτώ θέσεις είναι $\frac{1}{10}\left(\frac{9}{10}\right)^8$. Με την ίδια πιθανότητα προκύπτει δείγμα με ένα "και" σε οποιαδήποτε από τις εννέα θέσεις και

"όχι και" στις υπόλοιπες οκτώ. Άρα έχουμε συνολικά πιθανότητα $9\,\frac{1}{10}\,(\frac{9}{10})^8$ να προκύψει το ενδεχόμενο D με δεδομένο το ενδεχόμενο N:

$$P(D|N) = 9\,\frac{1}{10}\left(\frac{9}{10}\right)^8.$$

Με τον ίδιο τρόπο βρίσκουμε ότι η πιθανότητα να προκύψει το ενδεχόμενο D με δεδομένο το ενδεχόμενο M (δηλαδή με δεδομένο ότι το κείμενο είναι της Μηδουλίδου) είναι:

$$P(D|M) = 9\,\frac{2}{10}\left(\frac{8}{10}\right)^8.$$

Αυτό που ζητάμε είναι η πιθανότητα να είναι το κείμενο της Νταμά, με δεδομένο ότι έχει προκύψει δείγμα εννέα λέξεων με μόνο ένα "και", δηλαδή η πιθανότητα $P(N|D)$. Ο τύπος είναι:

$$P(N|D) = \frac{P(D|N)P(N)}{P(D|N)P(N) + P(D|M)P(M)} = \frac{9\,\frac{1}{10}\,(\frac{9}{10})^8\,\frac{6}{10}}{9\,\frac{1}{10}\,(\frac{9}{10})^8\,\frac{6}{10} + 9\,\frac{2}{10}\,(\frac{8}{10})^8\,\frac{4}{10}}.$$

3. *Σε μια ερώτηση πολλαπλής επιλογής υπάρχουν n επιλογές από τις οποίες μόνο μία είναι σωστή. Ο φοιτητής Αναξίμανδρος σημειώνει την σωστή επιλογή αν την γνωρίζει ή επιλέγει στην τύχη μία από τις n επιλογές. Υποθέτουμε ότι η πιθανότητα να γνωρίζει ο Αναξίμανδρος την σωστή απάντηση είναι p. Αν ο Αναξίμανδρος απάντησε σωστά, ποιά είναι η πιθανότητα να γνώριζε την σωστή επιλογή;*
   Γνωρίζουμε ότι το ενδεχόμενο Γ, δηλαδή "ο Αναξίμανδρος γνωρίζει την σωστή απάντηση", έχει πιθανότητα $P(\Gamma) = p$. Επομένως, το ενδεχόμενο ΔΓ, δηλαδή "ο Αναξίμανδρος δεν γνωρίζει την σωστή απάντηση", έχει πιθανότητα $P(\Delta\Gamma) = 1 - p$. Επίσης, γνωρίζουμε ότι, με δεδομένο το ενδεχόμενο Γ, το ενδεχόμενο Σ, δηλαδή "ο Αναξίμανδρος σημειώνει την σωστή απάντηση" είναι $P(\Sigma|\Gamma) = 1$. Ενώ, με δεδομένο το ενδεχόμενο ΔΓ, το ενδεχόμενο Σ, δηλαδή "ο Αναξίμανδρος σημειώνει την σωστή απάντηση" είναι $P(\Sigma|\Delta\Gamma) = \frac{1}{n}$ (διότι όταν δεν γνωρίζει την σωστή απάντηση επιλέγει στην τύχη μία από τις n απαντήσεις).
   Ψάχνουμε την πιθανότητα να γνωρίζει ο Αναξίμανδρος την σωστή απάντηση, με δεδομένο ότι σημειώνει την σωστή απάντηση, δηλαδή την $P(\Gamma|\Sigma)$. Ο τύπος είναι:

$$P(\Gamma|\Sigma) = \frac{P(\Sigma|\Gamma)P(\Gamma)}{P(\Sigma|\Gamma)P(\Gamma) + P(\Sigma|\Delta\Gamma)P(\Delta\Gamma)} = \frac{1\,p}{1\,p + \frac{1}{n}\,(1-p)}.$$

4. *Το ποσοστό των κατοίκων της νήσου Ευτυχία που πάσχει από την σοβαρή ασθένεια της κατάθλιψης είναι $0,01$. Ο προβληματισμένος Ευτυχιώτης ονόματι Χαμογελαστό Καλαμάρι υποβάλλεται σε δύο ανεξάρτητα τεστ για την ασθένεια καθένα από τα οποία κάνει ορθή διάγνωση με πιθανότητα $0,95$. Ποιά είναι η πιθανότητα να πάσχει το Χαμογελαστό Καλαμάρι αν (i) ένα τουλάχιστον τεστ είναι θετικό; (ii) και τα δύο τεστ είναι θετικά;*
   (i) Γνωρίζουμε ότι το ενδεχόμενο Π, δηλαδή "ο ΧΚ πάσχει", έχει πιθανότητα $P(\Pi) = \frac{1}{100}$. Επομένως, το ενδεχόμενο Υ, δηλαδή "ο ΧΚ είναι υγιής", έχει πιθανότητα $P(\Upsilon) = \frac{99}{100}$. Επίσης, γνωρίζουμε ότι για οποιοδήποτε από τα δύο τεστ, με δεδομένο το ενδεχόμενο Π, το ενδεχόμενο Θ, δηλαδή "το τεστ είναι θετικό", έχει πιθανότητα $P(\Theta|\Pi) = \frac{95}{100}$. Ενώ, με δεδομένο το ενδεχόμενο Π, το ενδεχόμενο A, δηλαδή "το τεστ είναι αρνητικό", έχει πιθανότητα $P(A|\Pi) = \frac{5}{100}$. Ομοίως, $P(\Theta|\Upsilon) = \frac{5}{100}$ και $P(A|\Upsilon) = \frac{95}{100}$.
   Για τα δύο ανεξάρτητα τεστ έχουμε τα εξής τέσσερα ενδεχόμενα:

$$\Theta\Theta \quad \Theta A \quad A\Theta \quad AA$$

Τότε η πιθανότητες αυτών των ενδεχομένων, με δεδομένο το ενδεχόμενο Π είναι

$$P(\Theta\Theta|\Pi) = \left(\frac{95}{100}\right)^2 \quad P(\Theta A|\Pi) = P(A\Theta|\Pi) = \frac{5}{100}\,\frac{95}{100} \quad P(AA|\Pi) = \left(\frac{5}{100}\right)^2$$

Ομοίως,

$$P(\Theta\Theta|\Upsilon) = \Big(\frac{5}{100}\Big)^2 \quad P(\Theta A|\Upsilon) = P(A\Theta|\Upsilon) = \frac{95}{100}\,\frac{5}{100} \quad P(AA|\Upsilon) = \Big(\frac{95}{100}\Big)^2$$

Από την πρώτη τετράδα έχουμε ότι, με δεδομένο το Π, το ενδεχόμενο $ET\Theta$, δηλαδή "ένα τουλάχιστον τεστ είναι θετικό", έχει πιθανότητα

$$P(ET\Theta|\Pi) = \Big(\frac{95}{100}\Big)^2 + 2\,\frac{5}{100}\,\frac{95}{100}.$$

Από την δεύτερη τετράδα έχουμε ότι, με δεδομένο το Υ, το ενδεχόμενο $ET\Theta$, δηλαδή "ένα τουλάχιστον τεστ είναι θετικό", έχει πιθανότητα

$$P(ET\Theta|\Upsilon) = \Big(\frac{5}{100}\Big)^2 + 2\,\frac{5}{100}\,\frac{95}{100}.$$

Ζητάμε την πιθανότητα του ενδεχομένου Π, δεδομένου του ενδεχομένου $ET\Theta$, δηλαδή την $P(\Pi|ET\Theta)$. Ο τύπος είναι

$$\begin{aligned} P(\Pi|ET\Theta) &= \frac{P(ET\Theta|\Pi)P(\Pi)}{P(ET\Theta|\Pi)P(\Pi) + P(ET\Theta|\Upsilon)P(\Upsilon)} \\ &= \frac{\big((\frac{95}{100})^2 + 2\,\frac{5}{100}\,\frac{95}{100}\big)\frac{1}{100}}{\big((\frac{95}{100})^2 + 2\,\frac{5}{100}\,\frac{95}{100}\big)\frac{1}{100} + \big((\frac{5}{100})^2 + 2\,\frac{5}{100}\,\frac{95}{100}\big)\frac{99}{100}}. \end{aligned}$$

(ii) Τώρα έχουμε

$$\begin{aligned} P(\Pi|\Theta\Theta) &= \frac{P(\Theta\Theta|\Pi)P(\Pi)}{P(\Theta\Theta|\Pi)P(\Pi) + P(\Theta\Theta|\Upsilon)P(\Upsilon)} \\ &= \frac{(\frac{95}{100})^2\,\frac{1}{100}}{(\frac{95}{100})^2\,\frac{1}{100} + (\frac{5}{100})^2\,\frac{99}{100}}. \end{aligned}$$

5. *Αν τα ενδεχόμενα A, B, C είναι ανεξάρτητα, αποδείξτε ότι τα ενδεχόμενα $A \cap B, C$ καθώς και τα ενδεχόμενα A^c, B^c, C^c είναι ανεξάρτητα.*

$$P((A\cap B)\cap C) = P(A\cap B\cap C) = P(A)P(B)P(C) = P(A\cap B)P(C).$$

Άρα τα $A \cap B, C$ είναι ανεξάρτητα.

$$\begin{aligned} P(A^c\cap B^c) &= P((A\cup B)^c) = 1 - P(A\cup B) = 1 - (P(A) + P(B) - P(A\cap B)) \\ &= 1 - P(A) - P(B) + P(A)P(B) = (1-P(A))(1-P(B)) \\ &= P(A^c)P(B^c). \end{aligned}$$

Ομοίως, έχουμε ότι $P(A^c\cap C^c) = P(A^c)P(C^c)$ και $P(B^c\cap C^c) = P(B^c)P(C^c)$.
Επίσης,

$$\begin{aligned} P(A^c\cap B^c\cap C^c) &= P((A\cup B\cup C)^c) = 1 - P(A\cup B\cup C) \\ &= 1 - (P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) \\ &\quad + P(A\cap B\cap C)) \\ &= 1 - P(A) - P(B) - P(C) + P(A)P(B) + P(A)P(C) + P(B)P(C) \\ &\quad - P(A)P(B)P(C) \\ &= (1-P(A))(1-P(B))(1-P(C)) = P(A^c)P(B^c)P(C^c). \end{aligned}$$

Άρα τα A^c, B^c, C^c είναι ανεξάρτητα.

6. *Έστω $A_1, A_2, \ldots$ ξένα ανά δύο ενδεχόμενα και ένα ενδεχόμενο $A \subseteq \bigcup_n A_n$. Αποδείξτε ότι για κάθε ενδεχόμενο B ισχύει*

$$P(B|A) = \sum_n P(B|A \cap A_n) P(A_n|A).$$

Αυτός είναι ο τύπος ολικής πιθανότητας για δεσμευμένες πιθανότητες.
Ειδικώτερα, αν $A = \bigcup_n A_n$, τότε

$$P(B|A) = \sum_n P(B|A_n)P(A_n) \Big/ \sum_n P(A_n).$$

Έχουμε ότι

$$\begin{aligned}
\sum_n P(B|A \cap A_n)P(A_n|A) &= \sum_n \frac{P(B \cap A \cap A_n)}{P(A \cap A_n)} \frac{P(A_n \cap A)}{P(A)} \\
&= \sum_n \frac{P(B \cap A \cap A_n)}{P(A)} = \frac{\sum_n P(B \cap A \cap A_n)}{P(A)} \\
&= \frac{P(\bigcup_n (B \cap A \cap A_n))}{P(A)} = \frac{P(B \cap A \cap \bigcup_n A_n)}{P(A)} \\
&= \frac{P(B \cap A)}{P(A)} = P(B|A).
\end{aligned}$$

Η τέταρτη ισότητα ισχύει διότι τα σύνολα $B \cap A \cap A_n$ είναι ανά δύο ξένα (επειδή τα A_n είναι ανά δύο ξένα). Η έκτη ισότητα ισχύει επειδή $A \subseteq \bigcup_n A_n$ και άρα $A \cap \bigcup_n A_n = A$. Αν $A = \bigcup_n A_n$, τότε $A \cap A_n = A_n$ και άρα ο n-οστός όρος στο άθροισμα του πρώτου τύπου είναι

$$\begin{aligned}
P(B|A \cap A_n)P(A_n|A) = P(B|A_n)P(A_n|A) &= P(B|A_n) \frac{P(A_n \cap A)}{P(A)} \\
&= \frac{P(B|A_n)P(A_n)}{P(A)}.
\end{aligned}$$

Άρα ο πρώτος τύπος γίνεται

$$\begin{aligned}
P(B|A) &= \sum_n P(B|A \cap A_n)P(A_n|A) = \sum_n \frac{P(B|A_n)P(A_n)}{P(A)} = \frac{\sum_n P(B|A_n)P(A_n)}{P(A)} \\
&= \frac{\sum_n P(B|A_n)P(A_n)}{\sum_n P(A_n)}.
\end{aligned}$$