Συμπλήρωμα στις Διακριτές Κατανομές

* Συμπληρώνουμε τις διαλέξεις που αφορούσαν στα τέσσερα παραδείγματα διακριτών

* Δίνονται ερμηνείες στα πλαίσια τυχαίων πειραμάτων κ' τυχαίων μεταβλητών

* Αποδεικνύουμε ένα οριακό θεώρημα

* Δίνουμε ένα παράδειγμα χρήσης της διωνυμικής κατανομής

Ερμηνεία των Διακριτών Κατανομών που εξετάσαμε.

Συμπλήρωμα στα παραδείγματα των διακριτών κατανομών: Ερμηνεία μέσω πειραμάτων τύχης, τυχαίων μεταβλητών κ' ορίων.

Στα παρακάτω θα δοθεί ερμηνεία των παραδειγμάτων διακριτών κατανομών μέσω πειραμάτων τύχης, εμφύτευσης των στοιχειωδών ενδεχομένων τους στο $\mathbb{R}$ μέσω τυχαίων μεταβλητών κ' ορίων ως προς το πλήθος των πειραμάτων.

1. Εκφυλισμένη κατανομή στο 0

Έστω πείραμα τύχης με μοναδικό (κ' συνεπώς βέβαιο) ενδεχόμενο α. Το σύνολο των στοιχειωδών ενδεχομένων είναι το $\Omega=\{\alpha\}$, η συλλογή των σύνθετων ενδεχομένων, $\Sigma_\Omega = \{\emptyset,\Omega\} = \{\emptyset,\{\alpha\}\}$ κ' η μοναδική (γιατί;) κατανομή που μπορεί να ορίζεται στο Σ_Ω είναι αυτή που αποδίδει $\mathbb{P}(\emptyset)=0$, $\mathbb{P}(\Omega)=\mathbb{P}(\{\alpha\})=1$.

Έστω τυχαία μεταβλητή $X:\Omega\to\mathbb{R}$ που ορίζεται ως

$X(\alpha)=0$. (γιατί είναι καλώς ορισμένη;). Τότε η X

μεταφέρει την $\mathbb{P}$ στο $\mathbb{R}$ αποδίδοντας την κατανομή στους

πραγματικούς που ορίζεται ως, $A \in \Sigma_{\mathbb{R}}$, $\mathbb{P}(A)=\begin{cases} 0, & 0 \notin A \\ 1, & 1 \in A \end{cases}$.

(δείξτε το!) Αυτή όμως είναι η εκφυλισμένη στο 0 (γιατί;).

Οπότε η εκφυλισμένη κατανομή προκύπτει από την παραπάνω "εμφύτευση" του στοιχειώδους ενδεχομένου του παραπάνω πειράματος στο $\mathbb{R}$ (αντιστοίχιση του α στο 0 μέσω της X).

Άσκηση: Αν είχαμε χρησιμοποιήσει την τυχαία μεταβλητή $Y:\Omega\to\mathbb{R}$, $Y(\alpha)=1$ ποια κατανομή θα είχε προκύψει στο $\mathbb{R}$; Γενικεύστε.

2. Bernoulli με παράμετρο $q\in(0,1)$ [Ber(q)]

Έστω τυχαίο πείραμα με δύο στοιχειώδη ενδεχόμενα α, β.

Αντιστοίχα με τα παραπάνω έχουμε $\Omega=\{\alpha,\beta\}$, $\Sigma_\Omega=\{\emptyset,\{\alpha\},\{\beta\},\Omega\}$

Για $q\in(0,1)$, έστω η κατανομή επί του Ω που ορίζεται από:

$\mathbb{P}(\emptyset)=0$, $\mathbb{P}(\{\alpha\})=q$, $\mathbb{P}(\{\beta\})=1-q$, $\mathbb{P}(\Omega)=1$ (την έχουμε ξαναδεί - γιατί είναι καλώς ορισμένη;) Έστω η τυχαία μεταβλητή

$X:\Omega\to\mathbb{R}$ που ορίζεται ως $X(\alpha)=1$, $X(\beta)=0$ (γιατί είναι καλώς ορισμένη;)

Η συγκεκριμένη X μεταφέρει την $\mathbb{P}$ στους πραγματικούς αριθμούς, αποκτώντας την κατανομή επί του $\mathbb{R}$ που ορίζεται ως: $A \in \Sigma_{\mathbb{R}}$, $\mathbb{P}(A) = \begin{cases} 0, & 0,1 \notin A \\ 1-q, & 0 \in A, 1 \notin A \\ q, & 0 \notin A, 1 \in A \\ 1, & 0,1 \in A \end{cases}$ (Δείξτε το!)

Αυτή όμως είναι η $Ber(q)$ (γιατί;).

3. Διωνυμική κατανομή με παραμέτρους $n \in \mathbb{N}^*$, $q \in (0,1)$ [$Bin(n,q)$]

Έστω ότι το τυχαίο πείραμα στο $\mathcal{A}$ επαναλαμβάνεται n φορές με "ανεξάρτητο τρόπο" (δηλ. το αποτέλεσμα όποιου πειράματος δεν επηρεάζει την έκβαση όποιου άλλου). Έστω $X_1, X_2, \ldots, X_n$ οι τυχαίες μεταβλητές που λειτουργούν σε κάθε ένα από τα πειράματα όπως η X στο $\mathcal{A}$ (δηλ. $\forall i=1,\ldots,n$, $X_i(\alpha)=1$, $X_i(\beta)=0$).

Έστω η τυχαία μεταβλητή $Y = \sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$.

Επειδή $\forall i=1,\ldots,n$ η X_i μπορεί να λάβει μόνο τις τιμές 0 ή 1, η Y μπορεί να λάβει τις τιμές $0, 1, 2, \ldots, n$ (γιατί;). Είναι δυνατόν να αποδειχθεί ότι $\mathbb{P}(Y=i) = \binom{n}{i} q^i (1-q)^{n-i}$ $\forall i=0,\ldots,n$.

(Προσπαθήστε το! Χρησιμοποιήστε την κατανομή στο $\mathcal{A}$ κ' την ανεξαρτησία)

Μα η παραπάνω πιθανότητα είναι αυτή που αποδίδει η $Bin(n,q)$ στο $\{i\}$. Επομένως η $Bin(n,q)$ προκύπτει από την $Ber(q)$

μέσω της ανεξάρτητης επανάληψης n φορές του αντίστοιχου τυχαίου πειράματος κ' της παραπάνω άθροισης.

Δ. Poisson με παράμετρο $\lambda>0$ [Poiss(λ)]

Έστω ότι στο 3 αφήνουμε το πλήθος των (ανεξάρτητων μεταξύ τους) πειραμάτων να τείνει στο $+\infty$ ($n \to +\infty$), αλλά για κάθε n, επιβάλλουμε στην Bin(n,q) που προκύπτει, το q να εξαρτάται από το n με τρόπο τέτοιο ώστε: καθώς το $n \to +\infty$, $q \to 0$ κ' $nq \to \lambda > 0$

(π.χ. έστω ότι για αρκετά μεγάλο n, $q := \lambda/n$).

[Κάτι τέτοιο θα ήταν εφικτό αν επιβάλλαμε στην Ber(q) που περιγράφει το n-ιοστό πείραμα, το q να εξαρτάται από το n με τον τρόπο που περιγράψαμε παραπάνω, $\forall n = 1, 2, \ldots$]

Πώς μπορούμε να αντιληφθούμε το $\lim_{n \to +\infty} \text{Bin}(n,q)$;

Θυμόμαστε ότι για την Bin(n,q), supp $= \{0, 1, \ldots, n\}$ επομένως είναι διαισθητικά εμφανές ότι όταν $n \to +\infty$, αν το όριο είναι καλώς ορισμένο θα έχει, supp $= \mathbb{N}$.

Επίσης για $i \leq n$ (← σταθερό), για την Bin(n,q), $\mathbb{P}(\{i\}) = \binom{n}{i} q^i (1-q)^{n-i}$

$$= \frac{n!}{i!\,(n-i)!} q^i (1-q)^{n-i} = \frac{(n-i)!\,(n-i+1)\cdot(n-i+2)\cdots n}{i!\,(n-i)!} \frac{(nq)^i}{n^i} (1-q)^{-i} \left(1 - \frac{nq}{n}\right)^n$$

$$= \frac{(n-i+1)\cdot(n-i+2)\cdots n}{n^i} \frac{1}{i!} (nq)^i (1-q)^{-i} \left(1-\frac{nq}{n}\right)^n$$

Παρατηρούμε ότι καθώς το $n \to +\infty$ $\underbrace{(n-i+1)(n-i+2)\cdots n}_{i \text{ όροι, } i \text{ σταθερό}}$ συμπεριφέρεται όπως n^i επομένως $\frac{(n-i+1)(n-i+2)\cdots n}{n^i} \to 1$.

Επίσης $(nq)^i \to \lambda^i$ αφού το $nq \to \lambda$,

$(1-q)^{-i} \to 1^{-i} = 1$ αφού $q \to 0$, (i σταθερό)

$\left(1 - \overset{\to\lambda}{nq}/n\right)^n \longrightarrow \exp(-\lambda)$ από τον ορισμό του e.

Τα παραπάνω μας λένε ότι για κάθε $i \in \mathbb{N}$, σταθερό

$$\lim_{n\to+\infty} \mathbb{P}(\{i\}) = \lim_{n\to+\infty} \binom{n}{i} q^i (1-q)^{n-i} = e^{-\lambda} \lambda^i / i!$$

Επομένως έχουμε αποκτήσει κατανομή (εφόσον είναι καλώς ορισμένη) με $\text{supp} = \mathbb{N}$, κ' $\mathbb{P}(\{i\}) = e^{-\lambda} \lambda^i / i!$, $i \in \mathbb{N}$.

Αυτή όμως είναι η $\text{Poiss}(\lambda)$.

Επομένως η $\text{Poiss}(\lambda)$ προκύπτει ως όριο της $\text{Bin}(n,q)$ καθώς $n \to +\infty$, $q \to 0$, $n.q \to \lambda$.

Το παραπάνω ονομάζεται Οριακό Θεώρημα Poisson

[Poisson Limit Theorem]

Παράδειγμα Χρήσης της Διωνυμικής Κατανομής.

- Έστω διαγώνισμα που αποτελείται από 30 ισό-βαθμες ερωτήσεις πολλαπλής επιλογής. Κάθε μία ερώτηση είναι ανεξάρτητη των υπολοίπων κ' έχει τέσσερις επιλογές εκ των οποίων μία είναι η σωστή.

Έστω φοιτητής που καλείται να απαντήσει στις ερωτήσεις (χωρίς εξωτερική βοήθεια). Αν ο φοιτητής δεν γνωρίζει την απάντηση σε καμμία από τις ερωτήσεις, η πιθανότητα να απαντήσει ορθά σε όποια από αυτές επιλέγοντας τυχαία απάντηση ισούται με 1/4. [Σε κάθε τέτοια ερώτηση έχουμε λοιπόν τυχαίο πείραμα - τυχαία επιλογή απάντησης - με το ενδεχόμενο αποτυχίας (λάθος απάντησης) να έχει πιθανότητα 3/4 κ' το ενδεχόμενο επιτυχίας (ορθή απάντηση) να έχει πιθανότητα 1/4.

$\Omega = \{\alpha, \beta\}$ α = σωστή απάντηση $X: \Omega \to \mathbb{R}$

β = λάθος απάντηση $X(\alpha) := 1, \; X(\beta) := 0$

$\mathbb{P}(\{\alpha\}) = 1/4$

$\mathbb{P}(\{\beta\}) = 3/4$ $\longrightarrow \mathrm{Ber}(q), \; q = 1/4$

Οπότε βάσει της προηγούμενης κατασκευής το τυχαίο αυτό πείραμα αναπαρίσταται από την $\mathrm{Ber}(1/4)$.]

Εξαιτίας της ανεξαρτησίας μεταξύ των ερωτήσεων και της ερμηνείας της $\mathrm{Bin}(n,q)$ θα έχουμε ότι η πιθανότητα ο τελείως αδιάβαστος φοιτητής να απαντήσει σωστά k ερωτήσεις $(k=0,1,\dots,30)$ θα ισούται $\mathbb{P}(\{k\}) = \binom{30}{k}(1/4)^k(3/4)^{30-k}$

Οπότε η τυχαιότητα της επίδοσης του φοιτητή στην εξέταση θα περιγράφεται από την

$$\mathrm{Bin}(30, 1/4).$$

Έτσι π.χ. $\mathbb{P}($ο εν λόγω φοιτητής να πάρει 5$)$

$\stackrel{*}{=} \mathbb{P}(\{15\}) = \binom{30}{15} (1/4)^{15} (3/4)^{15} \approx 0.0019$

(περίπου δύο τοις χιλίοις)

Άσκηση: Να βρεθεί η $\mathbb{P}($ο εν λόγω φοιτητής να περάσει το μάθημα$)$.

$\stackrel{*}{=} \mathbb{P}($ο εν λόγω φοιτητής να απαντήσει σωστά σε 15 ερωτήσεις$)$