Seminar 7

Variabile aleatoare discrete clasice.

In acest seminar ne intereseaza să aplicăm definiţia şi proprietăţile următoarelor distribuţii discrete clasice:

- Distribuţia Bernoulli;
- Distribuţia binomială;
- Distribuţia geometrică;
- Distribuţia Poisson;

7.1 Probleme rezolvate

1. (Run-uri de biţi) Având un şir de n biţi rezultaţi din simularea de n ori a variabilei aleatoare X de mai sus, numim run de biţi o succesiune de biţi identici în şir. De exemplu în şirul de biţi $0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0$, avem următoarele run-uri de biţi 1:

$$11, \quad 1, \quad 111$$

În analiza şirurilor de biţi aleatori folosiţi în probleme de securitate (criptografie) este foarte util numărul mediu de run-uri de biţi conţinuţi în şir.

Notăm cu N variabila aleatoare ce dă numărul de run-uri de biţi 1, în şirul de n biţi

$$b_1, b_2, \ldots, b_n,$$

obţinuţi din simularea (observarea) variabilelor aleatoare independente $X_1, X_2, \ldots, X_n$, Bernoulli distribuite, adică: $X_i = \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$

Practic simulăm aceeaşi variabilă aleatoare X, doar că îi asociem indicele $1, 2, 3, \ldots, n$ care indică al câtelea bit generăm.

Prin urmare fiecare bit b_i din şir este o "observaţie" asupra variabilei X_i, $i = \overline{1, n}$

Rezolvare:

Notăm cu Y_i variabila aleatoare ce ia valoarea 1 dacă în poziţia i a şirului de biţi începe un *run* de unu şi 0 în caz contrar. Astfel variabila aleatoare N, este suma variabilelor $Y_1, Y_2, \ldots Y_n$, $N = \sum_{i=1}^{n} Y_i$, iar numărul mediu de *run*-uri este $M(N) = \sum_{i=1}^{n} M(Y_i)$.

Dar conform definiţiei variabilelor Y_i avem că:

$$Y_1 = \begin{pmatrix} 1 & 0 \\ P(X_1 = 1) & P(X_1 = 0) \end{pmatrix}, \quad M(Y_1) = P(X_1 = 1) = p$$

adică un run de biţi începe din poziţia 1, dacă $X_1 = 1$. Pentru $i \geq 2$, un run de biţi începe din poziţia i, dacă bitul $b_i = 1$, dar bitul $b_{i-1} = 0$, adică pentru $i = \overline{2, n}$, $P(Y_i = 1) = P(X_{i-1} = 0, X_i = 1) = P(X_{i-1} = 0)P(X_i = 1) = (1-p)p$. Notând $\pi = (1-p)p$, variabila aleatoare Y_i are distribuţia de probabilitate:

$$Y_i = \begin{pmatrix} 1 & 0 \\ \pi & 1-\pi \end{pmatrix}$$

şi media $M(Y_i) = \pi = (1-p)p$. Prin urmare numărul mediu de *run*-uri de 1, în şirul de biţi $b_1, b_2, \ldots, b_n$, este $M(N) = p + (n-1)(1-p)p$.

Analiza statistică a *run*-urilor dintr-un şir de biţi se efectuează pentru a evalua caracterul aleator al acestora. Unul din testele folosite în criptografie este testul *run*, care numără *run*-urile de 0 şi 1 ditr-un şir de biţi produs de un generator de biţi aleatori. Dacă numărul *run*-urilor de 1,2,3,4,5 biţi şi respectiv ≥ 6 biţi, nu intră în intervale prescrise, şirul este considerat nesigur.

2. După ce un virus a pătruns în sistem, se verifică starea a 20 de fişiere. Ştiind că probabilitatea ca un fişier să fi fost afectat de virus este 0.2, independent de celelalte fişiere, să se determine probabilitatea ca cel mult cinci din cele 20 de fişiere să fi fost deteriorate ?

Rezolvare:

Variabila aleatoare X ce dă numărul fişierelor infectate din cele 20, are distribuţia de probabilitate Bin$(n = 20, p = 0.2)$. Probabilitatea cerută este

$$P(X \leq 5) = \sum_{k=0}^{5} C_{20}^{k} (0.2)^k (0.8)^{20-k}$$

3. Presupunem că există 10 sateliţi GPS (Global Positioning System) pe orbită. O unitate GPS de la sol este activă, dacă cel puţin 4 sateliţi GPS funcţionează (pot fi contactaţi). Ştiind că încercările de contactare ale celor 10 sateliţi sunt independente şi că probabilitatea ca GPS-ul de la sol să eşueze în contactarea oricărui satelit din cei 10, este aceeaşi şi egală cu $p = 0.75$, să se calculeze probabilitatea ca unitatea GPS de la sol să fie activă.

Rezolvare:

Încercările de contactare a celor $n = 10$ sateliţi GPS definesc un experiment Bernoulli, cu $p = 0.75$ (în acest context succesul nu are sensul uzual, ci $p = 0.75$ este probabilitatea ca GPS-ul de la sol să eşueze în contactarea unui satelit).

Variabila aleatoare X, ce dă numărul de sateliţi ce nu pot fi contactaţi de la sol, are distribuţia de probabilitate Bin$(n = 10, p = 0.75)$. Astfel probabilitatea ca unitatea GPS de la sol să funcţioneze este probabilitatea evenimentului $(X \leq 6)$, adică probabilitatea ca cel mult 6 sateliţi să nu poată fi contactaţi (deci cel puţin patru să poată fi contactaţi):

$$P(X \leq 6) = \sum_{k=0}^{6} C_{10}^{k}(0.75)^k(0.25)^{10-k}$$

4. Un student participă la un examen grilă cu 20 de probleme. Fiecare problemă are 4 răspunsuri posibile. Studentul cunoaşte răspunsul la 10 întrebări, iar la restul răspunde aleatoriu. Daca X este variabila aleatoare ce înregistrează numărul de răspunsuri corecte la acest test grilă, să se determine distribuţia de probabilitate acestei variabile. Care este probabilitatea ca studentul să răspunda corect la mai mult de 15 întrebări, din cele 20?
Rezolvare:

Notăm cu Y variabila aleatoare ce dă numărul de răspunsuri corecte la întrebările la care se răspunde aleator. Scorul final este dat de $X = Y + 10$. Vom determina distribuţia variabilei Y. Probabilitatea de succes (să se raspundă corect la o întrebare la care nu se ştie rezolvarea) este $p = \frac{1}{4}$. Atunci, Y se obţine prin 10 variabile Bernoulli independente, cu $p = 1/4$.
Deci, $Y \sim Bin(n = 10, p = 1/4)$. Avem că $p_k = P(Y = k) = C_{10}^{k}p^k(1-p)^{10-k}, k \in \{0, 1, \dots 10\}$. Pentru a determina distribuţia lui X, observăm că $D_X = \{10, 11, \dots, 20\}$. Vom analiza câte cazuri particulare:

$P(X = 10) = P(Y + 10 = 10) = P(Y = 0) = C_{10}^{0}p^0(1-p)^{10} = (\frac{3}{4})^{10}$
$P(X = 11) = P(Y + 10 = 11) = P(Y = 1) = C_{10}^{1}p^1(1-p)^9 = 10(\frac{1}{4})^1(\frac{3}{4})^9$

In general,

$$P(X = m) = P(Y + 10 = m) = P(Y = m - 10) = C_{10}^{m-10}p^{m-10}(1-p)^{20-m}, \forall m \in D_X$$

In plus, $P(X > 15) = P(X = 16) + P(X = 17) + \ldots + P(X = 20)$.

5. Se consideră jocul aruncării unui zar până când un număr mai mare decât 4 se obţine. Fie X variabila aleatoare ce înregistrează numărul de câte ori aruncăm zarul în acest caz. Să se determine $P(X = k), k = 1, 2, \dots$
Rezolvare:

La fiecare aruncare a zarului, probabilitate de succes $p = \frac{2}{6} = \frac{1}{3}$. Acest joc este un experiment Bernoulli repetat, până se obţine primul succes. Deci, $X \sim Geom(p = 1/3)$. Obţinem $P(X = k) = \frac{1}{3}(\frac{2}{3})^{k-1}, k = 1, 2, 3, \dots$

6. Aplicaţie Shazam identifică o melodie cu acurateţe de 95%.
Să se determine probabilitatea ca această aplicaţie să recunoască o melodie la a treia execuţie şi numărul mediu de încercări până la recunoaşterea melodiei (exclusiv)?
Rezolvare:

Fie Y variabila ce înregistrează numărul de încercări până la primul succes (recunoaşterea melodiei). Avem că $Y \sim Geom(p = 0.95)$. Vom calcula $P(Y = 3) = (1-p)^2p = (0.05)^2 0.95$.

Pentru a determina numărul mediu al încercărilor (eşecurilor), înaintea primului succes, vom defini variabila $W = Y - 1$. Avem că $M(W) = M(Y-1) = M(Y) - 1 = \frac{1}{p} - 1$.

7. (Coupon collector problem) O firmă lansează următorul concurs: în cutiile de cereale produse de această firmă se găsesc n cupoane diferite, ce pot fi colectate si trimise pentru un premiu. Se presupune că fiecare cupon din cutie este ales independent si uniform din cele n posibile şi că nu se fac schimburi de cupoane între participanţii la acest joc.
Să se determine numărul mediu de cutii de cereale ce trebuie cumpărate pentru a avea cel puţin unul din fiecare cupon?
Rezolvare: Fie X variabila aleatoare ce înregistrează numărul de cutii ce sunt cumpărate pentru a avea cel puţin unul din fiecare cupon. Ne interesează $M(X)$. Notăm cu X_i variabila aleatoare ce indică numărul de cutii cumpărate în timp ce ai deja $i-1$ cupoane diferite. Evident, $X = \sum\limits_{i=1}^{n} X_i$. Variabilele X_i sunt distribuite geometric: dacă ai $i-1$ cupoane diferite, atunci probabilitatea de succes (sa obţii unul nou) este $p = \frac{n-(i-1)}{n}$.
Deci, $M(X_i) = \frac{1}{p_i} = \frac{n}{n-i+1}$.
Din liniaritatea mediei, avem că $M(X) = \sum\limits_{i=1}^{n} M(X_i) = n \sum\limits_{i=1}^{n} \frac{1}{i}$.
Notăm $H_n = \sum\limits_{i=1}^{n} \frac{1}{i}$ şi se numeşte *număr armonic*. Se ştie că $H(n) = ln(n) + O(n)$.
De exemplu, pentru $n = 50$, trebuie să cumpărăm în medie 225 cutii.
Această problema are diverse aplicaţii în CS, pentru informatii suplimentare, see Probabilty - Computing, M. Mitzenmacher, Eli Upfal.

8. Numărul de mesaje email ce este primit într-o săptămână poate fi modelat de o variabilă Poisson cu media $\lambda = 0.2$ mesaje/min.

a) Să se determine probabilitatea de a nu primi nicun mesaj într-un interval de 5 minute.

b) Care este probabilitatea de a primi mai mult de 3 mesaje într-un interval de 10 minute?

Rezolvare:

a) Fie X variabila ce înregistrează numărul de mesaje primite în 5 minute. Atunci, $X \sim Pois(\lambda')$, unde λ' este media de mesaje primite in 5 minute. Se ştie că în medie se primesc $\lambda = 0.2$ mesaje/min, deci $\lambda' = 5(0.2) = 1$. Deci, $P(X = 0) = \frac{e^{-\lambda'}\lambda'^0}{0!} = \frac{1}{e}$.

b) Fie Y variabila ce înregistrează numărul de mesaje primite în 10 minute. Atunci, $Y \sim Pois(\lambda'')$, unde $\lambda^{''}$ este media de mesaje primite in 10 minute. Se ştie că în medie se primesc $\lambda = 0.2$ mesaje/min, deci $\lambda^{''} = 10(0.2) = 2$.
Avem $P(Y > 3) = 1 - P(Y \leq 3) = 1 - [P(Y = 0) + P(Y = 1) + P(Y = 2)]$.

7.2 Probleme propuse

9. Pe un site de cumpărături au intrat 10 vizitatori. Probabilitatea ca un vizitator să cumpere este 0.2.

a) Să se determine probabilitatea ca exact doi clienţi să cumpere de pe acest site.

b) Care este probabilitatea ca cel mult 4 clienţi să facă cumpărături?

10. Fie $X \sim Geom(p)$. Să se determine $M(X)$ şi $M(\frac{1}{2^X})$.

11. Probabilitatea ca un bit transmis printr-un canal de comunicaţie să fie primit cu eroare este 0.1. Se presupune că transmiterea bitilor se produce independent.

a) Să se determine probabilitatea ca primul bit transmis corect sa se produca la a cincea transmisie.

b) Care este numărul de biti transmisi până la primirea unui bit corect (inclusiv)?

12. Presupunem că apelurile telefonice la o staţie radio au probabilitate 0.02 de conectare. Se presupune că apelurile sunt independente.

a) Să se determine probabilitatea ca primul apel preluat să aibe loc la a zecea apelare.

b) Care este probabilitatea sa fie nevoie de mai mult de 5 apeluri pentru a te conecta?

c) Care este numărul mediu de apeluri necesare pentru a te putea conecta?

13. Pachetele de infomaţii ce se transmit între nodurile unei reţele pot interacţiona (intra in coliziune) cu probabilitate $p = 0.2$, indiferent de transmiterile precente.

a) Care este probabilitatea ca prima coliziune să aibă loc la a doua transmitere?

b) Care este probabilitatea ca la primele două transmiteri să nu se producă nici coliziune?

c) Să se determine numărul mediu de transmiteri făra coliziuni.

14. Fie $X \sim Pois(\lambda)$. Să se determine $M(X)$.

15. Numărul erorilor de tipar într-o carte urmează o distribuţie Poisson cu media $\lambda = 0.01$ erori/pag. Să se determine probabilitatea ca sa fie cel mult trei erori in 100 pagini.

16. Numărul mediu de cereri la o bază de date într-o perioadă de 15 minute este 10. Să se determine probabilitatea să se înregistreze exact 2 cereri în 3 minute.