VARIABILE ALEATOARE DISCRETE

Loredana Tănasie

November 15, 2023

Cuprins

1 Repartiţia Bernouli

Un tip simplu de variabilă aleatoare discretă este **variabila aleatoare Bernouli, care modelează efectuarea unui experiment în care poate apare unul din două rezultate posibile: numite succes, respectiv insucces**.

Aruncarea unei monede poate fi modelată printr-o variabilă aleatoare Bernoulli, spre exemplu dacă obţinerea stemei este succes.

Considerăm un spaţiu de selecţie cu două evenimente elementare $\mathcal{S}_1 = \{e_0, e_1\}$, pentru care

$$P\{e_0\}) = 1 - p \text{ şi } P(\{e_1\}) = p,$$

unde p este un număr oarecare din $[0, 1]$.

Realizarea evenimentului $\{e_1\}$, o considerăm obţinerea succesului şi codificăm cu valoarea 1, iar probabilitatea de realizare este p¸ iar insuccesul îl codificăm cu valoarea 0, iar probabilitatea este $1 - p$.
Reprezentarea variabilei aleatoare Bernoulli cu parametrul p (probabilitatea obţinerii succesului) are forma:

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Valoarea medie şi dispersia variabilei aleatoare Bernoulli cu parametrul p sunt date de

$$M(X) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$D^2(X) = (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p = p(1 - p)$$

2 Repartiţia uniformă

Dacă un experiment are n rezultate posibile egal posibile (notate $1, 2, \ldots, n$), atunci experimentul poate fi modelat printr-o **variabilă aleatoare uniformă** pe mulţimea $\{1, 2, \ldots, n\}$.
Variabila aleatoare uniformă pe mulţimea $\{1, 2, \ldots, n\}$ are repartiţia de forma

$$X : \begin{pmatrix} 1 & 2 & \ldots & n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix}$$

Valoarea medie şi dispersia variabilei aleatoare uniforme sunt date de

$$M(X) = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$$

$$\begin{aligned}
D^2(X) &= \sum_{i=1}^{n} \left(i - \frac{n+1}{2}\right)^2 \cdot \frac{1}{n} \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} i^2 - (n+1)\sum_{i=1}^{n} i + n\frac{(n+1)^2}{4}\right) \\
&= \frac{1}{n}\left(\frac{n(n+1)(2n+1)}{6} - (n+1)\frac{n(n+1)}{2} + n\frac{(n+1)^2}{4}\right) \\
&= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{2} + \frac{(n+1)^2}{4} \\
&= \frac{n^2-1}{12}.
\end{aligned}$$

3 Repartiţia binomială

Considerăm un spaţiu de selecţie cu două evenimente elementare $\mathcal{S}_1 = \{e_0, e_1\}$, pentru care

$$P(\{e_0\}) = 1 - p \text{ şi } P(\{e_1\}) = p,$$

unde p este un număr oarecare din $[0, 1]$.
Considerăm produsul cartezian

$$\mathcal{S} = \underbrace{\mathcal{S}_1 \times \mathcal{S}_1 \times ... \times \mathcal{S}_1}_{n \text{ ori}}$$

adică mulţimea elementelor de forma $(e_{i_1}, e_{i_2}, ..., e_{i_n})$ unde e_{i_j} este e_0 sau e_1.

Spaţiul de selecţie $\mathcal{S}$ conţine 2^n elemente. Probabilitatea fiecărui element din spaţiul de selecţie este:

$$P(\{(e_{i_1}, e_{i_2}, ..., e_{i_n})\}) = P_1(\{e_{i_1}\}) \cdot P_1(\{e_{i_2}\}) \cdot ... \cdot P_1(\{e_{i_n}\}).$$

Dacă A_k**,** $k = 0, 1, ..., n$ **este evenimentul care constă din acei** $(e_{i_1}, e_{i_2}, ..., e_{i_n})$ **care conţin** k **elemente de** e_1, atunci

$$P(A_k) = C_n^k \cdot p^k \cdot (1-p)^{n-k}.$$

Evenimentele A_0, A_1, ..., A_n sunt incompatibile şi reuniunea lor este evenimentul sigur. Variabila aleatoare X având repartiţia

$$X : \begin{pmatrix} 0 & 1 & ... & k & ... & n \\ C_n^0 p^0 (1-p)^n & C_n^1 p^1 (1-p)^{n-1} & ... & C_n^k p^k (1-p)^{n-k} & ... & C_n^n p^n (1-p)^0 \end{pmatrix}$$

se numeşte **variabilă binominală** şi se notează de obicei cu $B(n, p)$ (sau $X \sim B(n, p)$).

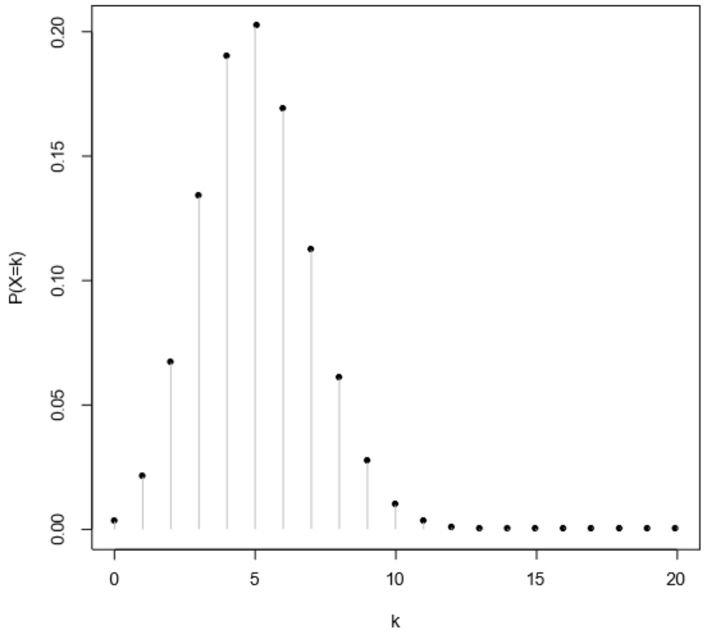
Mulţimea tuturor repartiţiilor binomiale poartă numele de **familia repartiţiilor binomiale**.

Valoarea medie şi dispersia unei variabile aleatoare $X \sim B(n, p)$ sunt: $M(X) = n \cdot p$ şi $D^2(X) = n \cdot p \cdot (1 - p)$.

Dacă $X_1 \sim B(n_1, p)$, $X_2 \sim B(n_2, p)$ şi X_1, X_2 sunt independente, $X_1 + X_2 \sim B(n_1 + n_2, p)$.

Funcţia de repartiţie a varibilei $X \sim B(n, p)$ este:

$$F(x) = \begin{cases} 0 & , x \leq 0 \\ (1-p)^n & , 0 < x \leq 1 \\ (1-p)^n + C_n^1 p (1-p)^{n-1} & , 1 < x \leq 2 \\ \dots \dots \dots & \\ (1-p)^n + C_n^1 p (1-p)^{n-1} + ... + C_n^k p^k (1-p)^{n-k} & , k < x \leq k+1 \\ \dots \dots \dots & \\ 1 & , x > n. \end{cases}$$



Funcția de probabilitate a unei variabile aleatoare *Bin*(20, 0.25)

4 Repartiţia Poisson

Adesea probabilităţile binominale $b(k;n,p) = C_n^k p^k (1-p)^{n-k}$ se calculează cu greutate. Tabelele construite pentru asemenea probabilităţi depind de doi parametri, n şi p.

$$C_n^k p^k q^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} =$$
$$\frac{\lambda^k}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}\cdot(1-\frac{\lambda}{n})^{n-k}$$

$$\lim_{n\to\infty}\frac{n(n-1)\dots(n-k+1)}{n^k} = 1, \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$$
$$\text{pentru } n\to\infty: \quad C_n^k p^k q^{n-k} \approx \frac{\lambda^k}{k!}e^{-\lambda}.$$

Teorema lui Poisson Dacă $n\cdot p_n \to \lambda$, atunci

$$b(k;n,p_n) \to \frac{\lambda^k}{k!}\,e^{-\lambda}.$$

Variabila aleatoare X care are funcţia de frecvenţă (funcţia de probabilitate)

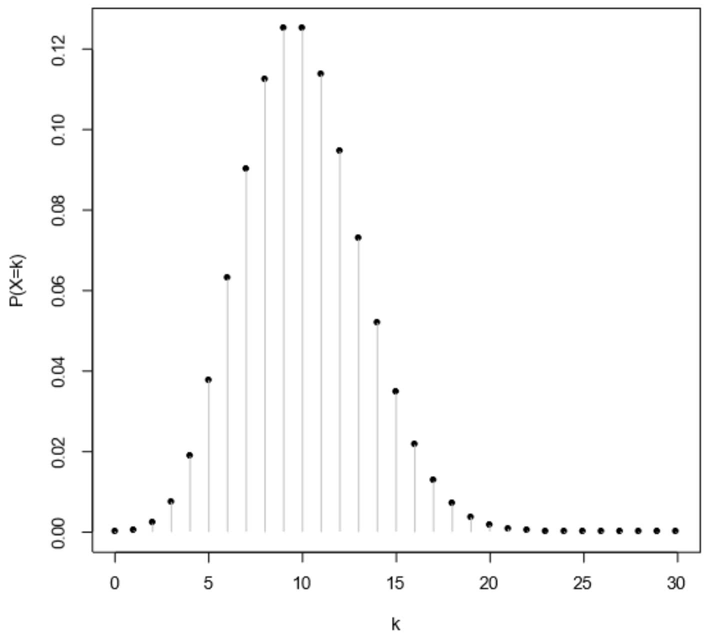
$$p(k,\lambda) = P(X=k) = \frac{\lambda^k}{k!}\,e^{-\lambda}\,, \quad k = 0,1,2,\dots$$

se numeşte **variabilă Poisson** cu parametrul λ şi se notează $X \sim Po(\lambda)$.

Repartiţia Poisson apare în multiple situaţii, de exemplu:

1. probabilităţile unui număr specificat de apeluri telefonice într-un anumit interval de timp;
2. probabilităţile unui număr specificat de defecte pe o unitate de lungime a unui fir;
3. probabilităţile unui număr specificat de defecte pe o unitate de arie a unei ţesături;
4. probabilităţile unui număr specificat de bacterii pe unitatea de volum într-o soluţie;
5. probabilităţile unui număr specificat de accidente în unitatea de timp.

Dacă X şi Y sunt variabile aleatoare independente repartizate Poisson de parametrii μ respectiv ν, atunci $X + Y \sim Po(\mu + \nu)$.



Funcția de probabilitate a unei variabile aleatoare *Po*(10)

5 Repartiţia multinominală

O urnă conţine $N = N_1 + N_2 + ... + N_k$ bile, dintre care: N_1 au culoarea 1, N_2 au culoarea 2, ..., N_k au culoarea k. Probabilitatea de a extrage o bilă oarecare din urnă este $\dfrac{1}{N}$. Probabilitatea de a extrage o bilă de culoare i din urnă este $p_i = \dfrac{N_i}{N}$. O urnă de acest fel se numeşte **urna lui Bernoulli cu mai multe stări**.

Fie A_i evenimentul care constă în extragerea unei bile de culoare i din urnă, $1 \leq i \leq k$.

Probabilitatea ca în n extrageri independente (fiecare dintre ele putând da naştere la unul din cele k evenimente A_1, A_2, ..., A_k) evenimentul A_i să se producă de n_i ori este

$$P_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!\, n_2! \ldots n_k!} \cdot p_1^{n_1}\, p_2^{n_2} \ldots p_k^{n_k}.$$

Un vector $(n_1, n_2, ..., n_k)$ având funcţia de frecvenţă

$$P_n(n_1, n_2, ..., n_k) = \frac{n!}{n_1!\, n_2! \ldots n_k!} \cdot p_1^{n_1}\, p_2^{n_2} \ldots p_k^{n_k}.$$

se numeşte multinomial.

Repartiţia corespunzătoare lui se numeşte **repartiţie multinomială** şi se notează cu $M(n; p_1, p_2, ..., p_n)$.

6 Repartiţia geometrică. Repartiţia binominală negativă

Considerăm situaţia în care numărul total de efectuări ale experienţei nu este dat dinainte.

Repetăm experienţa până obţinem r succese, acesta fiind fixat dinainte; numărul de insuccese (x) şi numărul total de repetări ale experienţei ($x + r$) sunt variabile.

Considerăm X variabila "numărul total de insuccese obţinute înainte de a se realiza succesul r".

Cazul $r = 1$**.**

Calculăm probabilitatea de a avea x insuccese înainte de **primul succes**.

Acest eveniment se poate realiza într-un singur mod: "trebuie să se obţină insuccese în primele x repetări ale experienţei şi apoi un succes în experienţa $x + 1$". Cum repetările experienţei sunt independente, avem

$$P(\underbrace{I, I, ..., I}_{x \text{ ori}}, S) = \underbrace{(1-p)\,(1-p)\,...\,(1-p)}_{x \text{ ori}}\, p\ ,$$

unde p este probabilitatea succesului. Prin urmare

$$f(x) = (1-p)^x \cdot p\ , \quad x = 0, 1, 2, ...$$

Variabila aleatoare X având funcţia de frecvenţă

$$f(x) = (1-p)^x \cdot p$$

se numeşte **variabilă geometrică**, iar $(x, f(x))$, $x = 0, 1, 2, ...$ se numeşte **repartiţie geometrică**.

Cazul $r \neq 1$**.**

Calculăm probablitatea de a obţine exact x insuccese înainte de **succesul** r.

Pentru a se realiza acest eveniment, trebuie să obţinem un succes în experienţa $r + x$ şi în cele $r + x - 1$ experienţe anterioare să obţinem $r - 1$ succese şi x insuccese. Probabilitatea de a obţine $r - 1$ succese în primele $r + x - 1$ experienţe este

$$C_{r+x-1}^{r-1} \cdot p^{r-1}\,(1-p)^x\ ,$$

unde probabilitatea de a obţine un succes în experienţa $r + x$ este p.

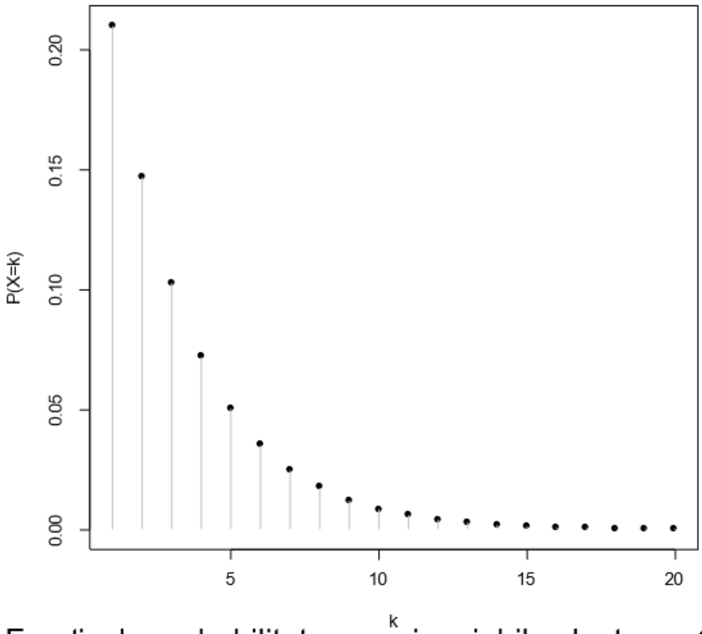
Deoarece experienţele sunt independente, probabilitatea căutată este

$$f(x) = C_{r+x-1}^{r-1} \cdot p^r\,(1-p)^x\ , \quad x = 0, 1, 2, ...$$

Variabila X având funcţia de frecvenţe

$$f(x) = C_{r+x-1}^{r-1} \cdot p^r\,(1-p)^x$$

se numeşte **variabilă binominală negativă**, iar $(x, f(x))$ **repartiţie binominală negativă**.



Funcția de probabilitate a unei variabile aleatoare *Geom*(0.3)

```
# R program to compute
# Negative Binomial Density

# Vector of x-values
x <- seq(0, 100, by = 1)

# Calling dnbinom() Function
y <- dnbinom(x, size = 10, prob = 0.5)

# Plot a graph
plot(y)
```

Output: