TEORIA DE LA PROBABILITAT

GM, FME, curs 2021–22

Tema 3: Variables aleatòries discretes

1. *(Distribució hipergeomètrica)* Una urna conté b boles blanques i g boles grogues. S'extreu una mostra de mida m sense replaçament i diem X_m al nombre de boles blanques que s'obtenen. Proveu que

$$p(X_m = k) = \frac{\binom{b}{k}\binom{g}{m-k}}{\binom{b+g}{m}},$$

i que aquesta expressió proporciona efectivament una distribució de probabilitat.

Proveu que si $g + b \to \infty$ amb $b/(g + b) \to p$, aleshores

$$\lim_{g+b\to\infty} p(X_m = k) = \binom{m}{k} p^k (1 - p)^{m-k}.$$

2. Proveu que, si X pren només valors enters no negatius, aleshores

$$\mathbb{E}[X] = \sum_{n\geq 1} p(X \geq n).$$

3. Una urna té b boles blanques i g boles grogues.

   (a) Es treuen boles fins que apareix la primera bola blanca. Proveu que el nombre mitjà d'extraccions és $(b + g + 1)/(b + 1)$.
   Indicació: Proveu que $\sum_{n=0}^{r} \binom{n+s}{s} = \binom{r+s+1}{s+1}$.

   (b) Es treuen boles fins que han sortit totes les d'un color. Quin és el nombre esperat de boles que queden a l'urna?

Solució: (b) $b/(g + 1) + g/(b + 1)$.

4. *(Problema del col·leccionista)* Un col·leccionista adquireix sobres tancats que contenen un de n cromos amb la mateixa probabilitat. Quin és el nombre esperat de sobres que ha d'adquirir per tenir la col·lecció completa?

Solució: $n \sum_{k=1}^{n} \frac{1}{k}$.

5. *(Hipergeomètrica negativa)* Una urna conté N boles de les quals b són blanques. Diem X el nombre de boles que s'han de treure fins a tenir-ne m de blanques. Proveu que la funció de probabilitat de X és

$$p(X = k) = \frac{b}{N}\binom{b-1}{m-1}\binom{N-b}{k-m}\Big/\binom{N-1}{k-1}.$$

Calculeu l'esperança $\mathbb{E}[X]$ de X.

Solució: $m(N+1)/(b+1)$.

6. Siguin X, Y dues variables aleatòries independents amb $p(X = k) = p(Y = k) = 1/2^k,\ k = 1, 2, \dots$. Trobeu:

   (a) $p(\min\{X, Y\} \leq k)$.

   (b) $p(Y > X)$.

   (c) $p(X = Y)$.

   (d) $p(X \geq nY)$ per $n \in \mathbb{N}$.

Solució: (a) $1 - 2^{-2k}$; (b) $1/3$; (c) $1/3$ (d) $2/(2^{n+1} - 1)$.

7. Siguin X, Y variables independents $B(1/2)$. Proveu que les variables $X + Y$ i $|X - Y|$ són incorrelades però no independents.

8. Siguin X i Y variables aleatòries de recompte. Siguin $G_X(z)$ i $G_Y(z)$ les funcions generadores de probabilitat de X i Y respectivament. Si $\alpha \in [0, 1]$, demostreu que $G_X(z)G_Y(z)$ i $\alpha G_X(z) + (1 - \alpha)G_Y(z)$ també defineixen funcions generadores de probabilitat. És també veritat per $G_X(\alpha z)/G_X(\alpha)$?

9. Sigui N el nombre de tirades d'una moneda amb probabilitat de cara p fins que surten n cares seguides. Doneu la funció generadora de probabilitats de N.

Solució: $G_N(z) = \frac{p^n z^n - p^{n+1} z^{n+1}}{1 - z + qp^n z^{n+1}}$

10. *(Fórmula de la variància iterada)* Es defineix la variància condicionada de Y donat X com $\mathbb{V}\mathrm{ar}[Y|X] = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$. Proveu que

$$\mathbb{V}\mathrm{ar}[Y] = \mathbb{E}[\mathbb{V}\mathrm{ar}[Y|X]] + \mathbb{V}\mathrm{ar}[\mathbb{E}[Y|X]].$$

11. Siguin X, Y variables independents Poisson amb paràmetres λ i μ, i sigui $Z = X + Y$. Proveu que, per n fix, $X|Z = n$ segueix una distribució binomial. Determineu-ne els paràmetres.

Solució: $\mathrm{Bin}(n, \lambda/(\lambda + \mu))$

12. Una gallina pon N ous en una setmana, on $N \sim \text{Po}(\lambda)$. Si cada ou té probabilitat $1/3$ d'acabar prosperant en pollet independentment dels altres, quina és la distribució del nombre M de pollets que s'obtenen dels ous d' una setmana? Calculeu $\mathbb{E}[N|M]$ i $\mathbb{E}[M|N]$.

    Solució: $M \sim \text{Po}(\lambda/3)$, $\mathbb{E}[M|N] = N/3$, $\mathbb{E}[N|M] = M + 2\lambda/3$.

13. Donada una v.a. X de recompte (és a dir, amb imatge $\mathbb{N} = \{0, 1, 2 \ldots\}$), es defineix Y (el *shift* de X) com la v.a. amb imatge $\mathbb{N}^+ = \{1, 2, \ldots, k, \ldots\}$, tal que $\forall k \in \mathbb{N}^+$ es compleix que $p\,(Y = k) = p\,(X = k - 1)$.

    (a) Si $G_X\,(z)$ és la fgp de X, trobeu-ne la de Y.

    (b) Si $Y \sim \text{Geom}\,(p)$, comproveu que hi ha una v.a. X, de forma que el shift de X és Y, direm que $X \sim \text{Geom}_0\,(p)$. Doneu-ne la fgp i calculeu-ne l'esperança i la variància.

    (c) Sigui $Z = \sum_{i=1}^{n} X_i$ on $n \in \mathbb{N}^+$ i les X_i són v.a. independents amb la mateixa distribució $\text{Geom}_0\,(p)$:

    i. Trobeu la funció de probabilitat de Z i la seva fgp.

    ii. Calculeu $\mathbb{E}\,[Z]$ i $\mathbb{V}\text{ar}\,[Z]$.

    iii. Dieu quina relació hi ha entre Z i una v.a de distribució $\text{BinN}\,(p, n)$.

14. *(Variables dobles de recompte)* Donada una v.a. doble de recompte $X = (X_1, X_2)$ on cada X_i és v.a. simple de recompte (és a dir, discreta amb imatge $\{0, 1, 2, \ldots, n, \ldots\}$), la fgp de X és $G_X\,(z_1, z_2) = \mathbb{E}\left[z_1^{X_1} \cdot z_2^{X_2}\right]$.

    (a) Comproveu:

    i. La fgp marginal de X_1 és $G_{X_1}\,(z_1) = G_X\,(z_1, 1)$, i anàlogament per X_2.

    ii. La fgp de $X_1 + X_2$ és $G_{(X_1+X_2)}\,(z) = G_X\,(z, z)$.

    iii. $G_X\,(z_1, z_2)$ és una funció analítica de (z_1, z_2) que com a mínim convergeix a $[-1, 1]^2$.

    iv. $p\,(X_1 = k_1, X_2 = k_2) = \frac{1}{k_1!k_2!} \frac{\partial^{(k_1+k_2)}}{\partial z_1^{k_1} \partial z_2^{k_2}} G_X\,(0, 0)$, amb el que $G_X\,(z_1, z_2)$ caracteritza $X = (X_1, X_2)$.

    v. $\text{Cov}\,(X_1, X_2) = \frac{\partial^2}{\partial z_1 \partial z_2} G_X\,(1, 1) - \frac{\partial}{\partial z_1} G_X\,(1, 1)\, \frac{\partial}{\partial z_2} G_X\,(1, 1)$.

    vi. X_1 i X_2 són independents $\Longleftrightarrow G_X\,(z_1, z_2) = G_{X_1}\,(z_1)\, G_{X_2}\,(z_2)$.

    (b) Calculeu:

    i. La fgp de $X = (X_1, X_2)$ i la $\text{Cov}\,(X_1, X_2)$, on $X_1 = B\,(p)$ i $X_2 = 1 - X_1$.

    ii. La fgp de $X = (X_1, X_2)$ i la $\text{Cov}\,(X_1, X_2)$, on $X_1 = \text{Bin}\,(n, p)$ i $X_2 = n - X_1$.

15. Siguin X, Y variables aleatòries independents per les quals $X|X+Y=n \sim \text{Bin}(n,p)$.
    (a) Proveu que $G_{X,Y}(t,s) = G_{X+Y}(pt+(1-p)s)$.
    (b) Proveu que X i Y segueixen una llei de Poisson.
    Indicació: podeu fer servir que les solucions de l'equació funcional $f(x+y) = f(x)f(y)$, $f(0)=1$ *i* f *contínua, són de la forma* e^{ax}.

16. *(Binomial subsampling)* Sigui X una variable de recompte. Un fet que es produeix molt sovint és que no arribem a poder comptar tots els individus o unitats de X. Per exemple:

    - El nombre d'individus d'una espècie en perill d'extinció en una zona determinada, en fer el recompte no tenim garanties que els hem vist tots.
    - El nombre de nous afectats d'una epidèmia durant una setmana, ja que és possible que no tots els afectats s'hagin reportat als serveis sanitaris.
    - El nombre d'agressions de gènere durant un any, ja que no totes es denuncien.

    El model del binomial subsampling planteja que cada unitat que hi ha de la variable X, té probabilitat p de ser comptabilitzada; o sigui el *p-thinning* de X, que denotarem per $p \circ X$, és la suma de Bernoulli's $B_k(p)$ independents (entre elles i de X), en la que el nombre de sumands és la variable X, per tant, $p \circ X = \sum_{k=1}^{X} B_k(p)$.

    Per $Y = p \circ X$, calculeu:

    (a) En general la fgp de Y, $G_Y(z)$.
    (b) La fgp de Y, $G_Y(z)$, en el cas en que $X \sim \text{Po}(\lambda)$. Comproveu que Y també és Poisson i trobeu-ne la seva λ.
    (c) En general $\mathbb{E}[Y]$ i $\mathbb{V}ar[Y]$.

17. *(Binomial subsampling, cont.)* Per la mateixa definició del binomial subsampling, els que no s'han pogut comptar $Z = X - Y$ és un thinning de X, $Z = (1-p) \circ X$. Considerem la variable doble (Y,Z) i volem saber si Y i Z estan relacionades o no, pel que calcularem la covariància $\text{Cov}(Y,Z)$ a partir de la fgp.

    Calculeu $G_{(Y,Z)}(t,s)$ i amb aquesta fgp trobeu $\text{Cov}(Y,Z)$ en els següents casos:

    (a) Si $X \sim \text{Unif}(0,n)$, calculeu $G_{(Y,Z)}(t,s)$ i $\text{Cov}(Y,Z)$. Caracteritzeu quan $\text{Cov}(Y,Z) = 0$.
    (b) Si $X \sim \text{Geom}(q)$, calculeu $G_{(Y,Z)}(t,s)$ i $\text{Cov}(Y,Z)$. Caracteritzeu quan $\text{Cov}(Y,Z) = 0$.
    (c) Si $X \sim \text{Geom}(q) - 1$, calculeu $G_{(Y,Z)}(t,s)$ i $\text{Cov}(Y,Z)$. Caracteritzeu quan $\text{Cov}(Y,Z) = 0$.

18. Siguin X i Y variables aleatòries de recompte. És cert que si $G_{X+Y}(z) = G_X(z)G_Y(z)$ aleshores X i Y són independents?

19. *(Random stopped sums)* Suposem que N, el nombre d'accidents de moto amb ferits que hi ha (lloc fixat i període fixat), és $N \sim \text{Po}(\lambda)$, i que el nombre de ferits en un accident de moto (amb ferits) pot ser 1 o 2 amb probabilitats p_1 i $p_2 = 1 - p_1$.

    (a) Trobeu la fgp de T (nombre total de ferits en accident de moto). La distribució de T s'anomena de Hermite.
    Calculeu-ne $p\,(T = k)$ per $k \leq 3$.

    (b) Calculeu l'esperança i la variància de T.

    (c) Si T_2 és el nombre total de ferits en accidents amb dos ferits i $T_1 = T - T_2$ la resta de ferits, són T_1 i T_2 independents? Quines distribucions tenen?

    (d) La distribució de Hermite és tancada per la suma (convolució)? És a dir, si $X_1 \sim \text{Hermite}\,(\lambda_1, p_{1,1})$ i $X_2 \sim \text{Hermite}\,(\lambda_2, p_{1,2})$ independents, aleshores $X_1 + X_2$ és Hermite? En cas afirmatiu, quins són els paràmetres?

20. *(Distribució logarítmica)*

    (a) Una variable aleatòria de recompte X té distribució logarítmica si la seva funció generadora de probabilitats té l'expressió $G_X\,(z) = C \cdot \log\,(1 - pz)$ amb $p \in (0, 1)$.

    i. Obteniu el valor de C per a que $G_X\,(z)$ sigui realment una fgp.

    ii. Calculeu la funció de probabilitat de X.

    iii. Calculeu l'esperança i la variància de X.

    (b) Sigui $Y = \sum_{i=1}^{N} X_i$ on les X_i son v.a. independents amb la mateixa distribució logarítmica, i $N \sim \text{Po}\,(\lambda)$ independent de les X_i. (Aclariment: $Y = 0$ si $N = 0$.)

    i. Comproveu que la fgp de Y es pot escriure de la forma $G_Y\,(z) = \left(\frac{1-q}{1-q\cdot z}\right)^r$. Doneu les expressions de r i q.

    ii. Calculeu la funció de probabilitat de Y.

    iii. Calculeu l'esperança i la variància de Y.

    iv. Y té distribució binomial negativa? r és enter o real?

21. Demostreu que en un arbre de Galton-Watson amb $\mu = \mathbb{E}[Z_1]$ and $\sigma^2 = \mathbb{V}\text{ar}[Z_1]$ es compleix que

$$\mu^n = \mathbb{E}[Z_n],\ \mathbb{V}\text{ar}[Z_n] = \begin{cases} n\sigma^2, & \mu = 1, \\ \sigma^2(\mu^n - 1)\mu^{n-1}(\mu - 1)^{-1}, & \mu \neq 1. \end{cases}$$

22. Suposeu que en un arbre de Galton-Watson $\mathbb{E}[z^X] = (2-z)^{-1}$, i $Z_0 = 1$. Sigui V_r el nombre total de generacions de amb r elements. Demostreu que $\mathbb{E}[V_1] = \frac{\pi^2}{6}$.

23. Demostreu que la funció generadora de probabilitat $H_n(z)$ del nombre total d'individus en les primeres n generacions d'un arbre de Galton-Watson satisfà l'equació $H_n(z) = zG(H_{n-1}(z))$, on $G(z)$ és la funció generadora de probabilitat de Z_1.

24. Considereu un arbre de Galton-Watson amb X una variable aleatòria del nombre de descendents de cada individu tal que $\mu = \mathbb{E}[X] > 1$. Considereu Ext l'esdeveniment que la població s'extingeixi eventualment en alguna generació i sigui $\eta = p(Ext) \in (0,1)$. Considereu $\hat{X}$ la variable aleatòria de recompte definida per tota $k \geq 0$

$$p(\hat{X} = k) = \eta^{k-1} p(X = k).$$

   (a) Demostreu que $\hat{X}$ és efectivament una distribució de probabilitat.

   (b) Sigui Z_n el nombre d'individus a la generació n-èsima. Proveu que $(Z_n | Ext)$ és un arbre de Galton-Watson on el nombre de descendents de cada individu segueix una distribució $\hat{X}$.