1 Osservazione su alcune v.a. discrete

Tra le v.a. discrete, abbiamo finora visto:

1. V.a. Uniforme sull'insieme $\{x_1, x_2, \ldots, x_n\}$; si denota con $X \sim Uni(x_1, x_2, \ldots, x_n)$.
X assume valori in $\{x_1, x_2, \ldots, x_n\}$ con uguali probabilità, ovvero $p_k = P(X = x_k) = 1/n, \; k = 1, 2, \ldots, n$.
Esempio: $X = \#$ che esce, lanciando un dado non truccato; in tal caso, $X \in \{1, 2, 3, 4, 5, 6\}$ e $p_k = 1/6, \; k = 1, 2, \ldots, 6$.

2.1 V.a. di Bernoulli di parametro $p \in (0, 1)$; si denota con $X \sim B(1, p)$.
X assume valori in $\{0, 1\}$ con probabilità:

$$p_0 = P(X = 0) = 1 - p, \; p_1 = P(X = 1) = p.$$

Esempio: si lancia una sola volta una moneta, per la quale la probabilità di uscita di Testa è $p \in (0, 1)$; se si scrive 1 sulla faccia Testa e 0 su quella Croce, $X \in \{0, 1\}$ è il risultato del lancio.

2.2 V.a. Binomiale di parametri $(n, p), \; p \in (0, 1)$; si denota con $X \sim B(n, p)$.
X assume valori in $\{0, 1, \ldots, n\}$ con probabilità:

$$p_k = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \; k = 0, 1, \ldots, n.$$

Esempio 1: $X = \#$ volte che esce Testa, lanciando n volte una moneta, per la quale la probabilità di uscita di Testa in ogni lancio è $p \in (0, 1)$; i risultati dei successivi lanci sono **indipendenti** tra loro.
Esempio 2: $X = \#$ di palline rosse estratte, in n estrazioni **con rimpiazzo**, da un'urna contenente r palline rosse e b bianche; $p = r/(b + r)$ è la probabilità di estrarre una pallina rossa, ad ogni estrazione; i risultati delle successive estrazioni sono **indipendenti** tra loro.
Si può scrivere $X = X_1 + X_2 + \cdots + X_n$, dove le v.a. $X_i \sim B(1, p), \; i = 1, 2, \ldots, n$, e le v.a. X_i sono indipendenti; X_i vale 1 se all'i-esimo lancio della moneta (o i-esima estrazione di una pallina, **con rimpiazzo**) esce Testa (o una pallina rossa), X_i vale 0, altrimenti.

3. V.a. Ipergeometrica di parametri (r, b, n); si denota con $X \sim Hyper(r, b, n)$.
X assume valori interi non negativi k tali che $n \leq b + r$ e $\max\{0, n - b\} \leq k \leq \min\{r, n\}$.
Si ha:

$$p_k = P(X = k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{b+r}{n}}$$

Esempio: X conta il numero di palline rosse estrattte in n estrazioni **senza rimpiazzo** da un'urna contenente r palline rosse e b bianche; i risultati delle successive estrazioni **non** sono indipendenti tra loro.
Si può ancora scrivere $X = X_1 + X_2 + \cdots + X_n$, dove le v.a. $X_i \sim B(1, p), \; i = 1, 2, \ldots, n$, però le v.a X_i **non** sono indipendenti; X_i vale 1 se all'i-esima estrazione di una pallina, **senza rimpiazzo**, esce una pallina rossa, X_i vale 0, altrimenti.

4. V.a. Geometrica modificata di parametro $p \in (0,1)$ (Istante di primo successo).
T assume valori $k \in \{1, 2, \dots, \}$. Si ha:

$$p_k = P(T = k) = p(1-p)^{k-1}, \; k = 1, 2, \dots$$

Esempio: $T =$ istante di primo successo, in una sequenza di prove indipendenti e Bernoulliane, in ciascuna delle quali la probabilità del successo è $p \in (0,1)$.
T gode della proprietà di mancanza di memoria:

$$P(T > n + m | T > n) = P(T > m),$$

ed è l'unica v.a. discreta, che assume valori interi positivi, a godere di tale proprietà.

5. V.a. Geometrica di parametro $p \in (0,1)$; si denota con $X \sim Geom(p)$.
X assume valori $k \in \{0, 1, 2, \dots, \}$. Si ha:

$$p_k = P(T = k) = p(1-p)^{k}, \; k = 0, 1, 2, \dots$$

Esempio: X rappresenta il numero degli insuccessi che precedono il primo successo, in una sequenza di prove indipendenti e Bernoulliane, in ciascuna delle quali la probabilità del successo è $p \in (0,1)$. Risulta $X = T - 1$, dove T è l'istante di primo successo. Anche X gode di una proprietà di mancanza di memoria, che si scrive ora:

$$P(X = n + m | X \geq n) = P(X = m).$$

X è l'unica v.a. discreta, che assume valori interi non negativi, a godere di tale proprietà.

SVOLGERE GLI ESERCIZI 1.12, 1.13, 1.18, 1.19, 1.20, 1.22, 1.29

$\triangleright$ **Esercizio 1.12**
Un ubriaco ha in tasca un mazzo di 8 chiavi tra cui vi è la sua chiave di casa. Giunto sulla soglia, egli cerca di aprire la porta della sua abitazione con una delle due seguenti procedure:
(i) tenta con una chiave scelta a caso; se la chiave non è quella giusta, la ripone nel mazzo e ne sceglie ancora una a caso e così via fino ad individuare la chiave giusta. Calcolare la probabilità che la chiave giusta venga scelta al primo, al secondo e, in generale, al $k-$esimo tentativo.
(ii) L'ubriaco, per un improvviso sprazzo di lucidità, cerca la chiave giusta, provando le chiavi a caso una dopo l'altra, ma senza rimettere nel mazzo le chiavi già provate (che non aprono). Calcolare la probabilità che la chiave giusta venga scelta al primo, al secondo e, in generale, al $k-$esimo tentativo.

Soluzione
(i) Il tentativo al quale l'ubriaco trova la chiave giusta è l'istante T di primo successo, in una successione di prove Bernoulliane e indipendenti, in cui la probabilità del successo in ciascuna prova è $p = 1/8$. Dunque:

$$P(T = k) = p(1-p)^{k-1} \ k = 1, 2, \ldots$$

e quindi:

$$P(T = 1) = \frac{1}{8} = 0.125, \ P(T = 2) = \frac{1}{8} \cdot \left(\frac{7}{8}\right)^{2-1} = \frac{7}{64} = 0.109 \ .$$

(ii) Denotiamo ancora con T il numero dei tentativi necessari per trovare la chiave giusta; siano: A l'evento "si sono ottenuti 0 successi nelle prime $k-1$ estrazioni senza rimpiazzo, tra 8 oggetti, dei quali uno solo corrisponde al successo"; B l'evento che la $k-$esima estrazione sia un successo. Si ha:

$$P(T = k) = P(A \cap B) = P(B|A)P(A).$$

Usando lo schema ipergeometrico, si ottiene:

$$P(A) = \frac{\binom{1}{0}\binom{7}{k-1}}{\binom{8}{k-1}} \ .$$

Inoltre:

$$P(B|A) = \frac{1}{8-(k-1)} \ .$$

Quindi, risulta:

$$P(T=k)=\frac{\binom{1}{0}\binom{7}{k-1}}{\binom{8}{k-1}}\cdot\frac{1}{8-(k-1)}=$$

$$=\frac{\binom{7}{k-1}}{\binom{8}{k-1}}\cdot\frac{1}{8-(k-1)}=\ldots=\frac{1}{8}\ ,\ k=1,2,\ldots,8.$$

Pertanto T ha distribuzione uniforme in $\{1,2,\ldots,8\}$.

▷ **Esercizio 1.13**
Sia X una v.a. a valori in $\mathbf{Z}$ tale che $P(X=X^3)=1$. Trovare la densità discreta di X, sapendo che $P(X=-1)=P(X=1)=p$. Se Y è un'altra v.a., indipendente da X, che assume valori $-1,1$ con probabilità $q,1-q$, calcolare la legge di $Z=X+Y$.

Soluzione
Siccome $x^3=x$ se e soltanto se $x=-1,0,1$, la v.a. X può assumere solo i valori $-1,0,1$.
Sappiamo già che $P(X=-1)=P(X=1)=p$, pertanto $P(X=0)=1-2p$.
La v.a. $Z=X+Y$ assume valori $k\in\{-2,-1,0,1,2\}$ e si ha:

$$P(Z=k)=P(X+Y=k)=\sum_{n=-1}^{1}P(X=n,Y=k-n);$$

per l'indipendenza di X e Y, tale formula diventa:

$$P(Z=k)=P(X+Y=k)=\sum_{n=-1}^{1}P(X=n)P(Y=k-n).$$

Dunque
per $k=-2:$

$$P(Z=-2)=P(X=-1)P(Y=-1)+$$
$$+P(X=0)P(Y=-2)+P(X=1)P(Y=-3)=$$
$$=pq+(1-2p)\cdot 0+p\cdot 0=pq;$$

per $k=-1:$

$$P(Z=-1)=P(X=-1)P(Y=0)+$$
$$+P(X=0)P(Y=-1)+P(X=1)P(Y=-2)=$$
$$=p\cdot 0+(1-2p)q+p\cdot 0=(1-2p)q;$$

per $k=0:$

$$P(Z=0)=P(X=-1)P(Y=1)+$$

$$+P(X=0)P(Y=0)+P(X=1)P(Y=-1)=$$
$$=p(1-q)+(1-2p)\cdot 0+pq=p;$$

per $k=1$:

$$P(Z=1)=P(X=-1)P(Y=2)+$$
$$+P(X=0)P(Y=1)+P(X=1)P(Y=0)=$$
$$=p\cdot 0+(1-2p)(1-q)+p\cdot 0=1-q-2p+2pq;$$

per $k=2$:

$$P(Z=2)=P(X=-1)P(Y=3)+P(X=0)P(Y=2)+$$
$$+P(X=1)P(Y=1)=p\cdot 0+(1-2p)\cdot 0+p(1-q)=p(1-q).$$

E' immediato verificare che, come deve essere, $P(Z=-2)+P(Z=-1)+P(Z=0)+P(Z=1)+P(Z=2)=1$.

▷ **Esercizio 1.18**
Siano X e Y v.a. indipendenti con distribuzioni uniformi su $\{1,2,\ldots,n\}$ e $\{1,2,\ldots,n,\ldots,2n\}$, rispettivamente. Trovare la legge di $Z=\max(X,Y)$; calcolarne inoltre la media.

Soluzione
La v.a. $Z=\max(X,Y)\in\{1,2,\ldots,n,n+1,\ldots,2n\}$. Se $1\le k\le n$, procedendo come nell'Esercizio 1.17, si ha:

$$P(Z=k)=P(X=k)P(Y=k)+P(X<k)P(Y=k)+P(X=k)P(Y<k)$$
$$\frac{1}{n}\cdot\frac{1}{2n}+\frac{k-1}{n}\cdot\frac{1}{2n}+\frac{1}{n}\cdot\frac{k-1}{2n}=\frac{2k-1}{2n^2}\ .$$

Se $n+1\le k\le 2n$, si ha invece:

$$P(Z=k)=P(Y=k)=\frac{1}{2n}\ ,$$

poiché in tal caso è $X<Y$. Pertanto:

$$E(Z)=\sum_{k=1}^{2n}kP(Z=k)=\sum_{k=1}^{n}k\frac{2k-1}{2n^2}+\sum_{k=n+1}^{2n}\frac{k}{2n}=$$
$$=\frac{1}{2n^2}\sum_{k=1}^{n}(2k^2-k)+\frac{1}{2n}\left[\sum_{k=1}^{2n}k-\sum_{k=1}^{n}k\right];$$

utilizzando le formule $(*)$ e $(**)$ dell'Esercizio 1.16, dopo alcuni calcoli e semplificazioni, si ottiene infine

$$E(Z)=\frac{13n^2+6n-1}{12n}\ .$$

Anche la varianza di Z si può calcolare in modo simile, però occorre conoscere la formula per la somma dei cubi dei primi n interi.

$\triangleright$ **Esercizio 1.19**
Con riferimento all'esercizio precedente, sia $U = \min(X, Y)$. Trovare la legge di U.

Soluzione
Ora $U = \min(X, Y) \in \{1, 2, \dots, n\}$ e risulta, per $k = 1, 2, \dots, n$:

$$P(U = k) = P(X = k)P(Y = k) + P(X = k)P(Y > k) + P(Y = k)P(X > k) =$$

$$\frac{1}{n} \cdot \frac{1}{2n} + \frac{1}{n}\left(1 - \frac{k}{2n}\right) + \frac{1}{2n}\left(1 - \frac{k}{n}\right) = \frac{1 + 3n - 2k}{2n^2} \ .$$

La media di U si potrebbe ottenere, con calcoli un po' laboriosi, utilizzando le formule per la somma dei primi n interi e dei quadrati dei primi n interi, viste negli esercizi precedenti.

$\triangleright$ **Esercizio 1.20**
Due dadi equilibrati vengono lanciati separatamente più volte. Indichiamo con X il numero di lanci necessario ad ottenere 3 gettando il primo dado, e con Y il numero dei lanci necessario a ottenere 2 oppure 5 lanciando il secondo.
(i) Qual è la legge di X? e di Y? Qual è $E(X)$? e $E(Y)$?
(ii) Trovare la densità discreta di $Z = \max(X, Y)$ e $E(Z)$.
(iii) calcolare $P(X \geq Y)$.

Soluzione
(i) X è l'istante di primo successo in uno schema di prove indipendenti e Bernoulliane, in cui la probabilità del successo in ciascuna prova è $p = 1/6$. Analogamente, Y è il tempo di primo successo in uno schema di prove indipendenti e Bernoulliane, in cui la probabilità del successo in ciascuna prova è $p' = 2/6 = 1/3$. Le v.a. X e Y hanno dunque legge geometrica modificata di parametro p e p', rispettivamente, ovvero risulta, per $k = 1, 2, \dots$:

$$P(X = k) = p(1 - p)^{k-1}, \quad P(Y = k) = p'(1 - p')^{k-1}.$$

Poiché X e Y sono v.a. positive, per calcolare $E(X)$ e $E(Y)$ possiamo utilizzare la formula (vista nell' Esercizio 1.15):

$$E(X) = \sum_{k=0}^{\infty} P(X > k)$$

A questo scopo, calcoliamo prima $P(X \leq k)$; si ha:

$$P(X \leq k) = \sum_{h=1}^{k} p(1 - p)^{h-1} = p \sum_{m=0}^{k-1} (1 - p)^m =$$

$$= p\left(\frac{1 - (1 - p)^k}{1 - (1 - p)}\right) = p \cdot \frac{1 - (1 - p)^k}{p} = 1 - (1 - p)^k.$$

Pertanto:
$$P(X > k) = 1 - P(X \leq k) = (1-p)^k.$$
Sostituendo nella formula di sopra, otteniamo:
$$E(X) = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p} = \frac{1}{1/6} = 6 \ .$$
Analogamente, risulta:
$$E(Y) = \frac{1}{p'} = \frac{1}{1/3} = 3 \ .$$
Dalla formula trovata, possiamo riottenere la media di una v.a. geometrica G di parametro p, per la quale cioè vale:
$$P(G = k) = p(1-p)^k, \ k = 0, 1, 2, \ldots$$
Infatti, possiamo porre $G = X - 1$ e quindi $E(G) = E(X) - 1 = 1/p - 1 = (1-p)/p$.
(ii) Se $Z = \max(X, Y)$, si ha, per $k = 1, 2, \ldots$:
$$P(Z \leq k) = P(X \leq k, Y \leq k)$$
che, per l'indipendenza di X e Y, e utilizzando i risultati trovati sopra vale:
$$P(X \leq k)P(Y \leq k) = [1 - (1-p)^k][1 - (1-p')^k] \ .$$
Allora:
$$P(Z = k) = P(Z \leq k) - P(Z \leq k-1) =$$
$$[1 - (1-p)^k][1 - (1-p')^k] - [1 - (1-p)^{k-1}][1 - (1-p')^{k-1}] \ .$$
Sostituendo i valori $p = 1/6$ e $p' = 1/3$, si ottiene $(k = 1, 2, \ldots)$:
$$P(Z = k) = \left[1 - \left(\frac{5}{6}\right)^k\right]\left[1 - \left(\frac{2}{3}\right)^k\right] - \left[1 - \left(\frac{5}{6}\right)^{k-1}\right]\left[1 - \left(\frac{2}{3}\right)^{k-1}\right];$$
svolgendo i calcoli, si ottiene infine:
$$P(Z = k) = \frac{1}{3}\left(\frac{2}{3}\right)^{k-1} + \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} - \frac{4}{9}\left(\frac{5}{9}\right)^{k-1} \ .$$
Si ha quindi:
$$E(Z) = \sum_{k=1}^{\infty} kP(Z = k) =$$
$$= \sum_{k=1}^{\infty} k \cdot \frac{1}{3}\left(\frac{2}{3}\right)^{k-1} + \sum_{k=1}^{\infty} k \cdot \frac{1}{6}\left(\frac{5}{6}\right)^{k-1} - \sum_{k=1}^{\infty} k \cdot \frac{4}{9}\left(\frac{5}{9}\right)^{k-1} \ .$$

Le tre serie scritte valgono, rispettivamente, la media di Y, la media di X e la media di un *istante di primo successo* relativo ad una serie di prove indipendenti e Bernoulliane di parametro $p'' = 4/9$. Dunque, per quanto visto sopra, si ha:

$$E(Z) = \frac{1}{1/3} + \frac{1}{1/6} - \frac{1}{4/9} = \frac{27}{4} \ .$$

(iii) Si ha:

$$P(X \geq Y) = \sum_{k=1}^{\infty} P(X \geq k, Y = k)$$

$$= \sum_{k=1}^{\infty} P(X \geq k) P(Y = k)$$

$$= \sum_{k=1}^{\infty} P(X > k-1) P(Y = k);$$

ricordando che per l'istante di primo successo T, cioè una v.a. Geometrica modificata di parametro p, vale che $P(T > h) = (1-p)^h, \ h = 1, 2, \dots$, si ha quindi:

$$P(X \geq Y) = \sum_{k=1}^{\infty} (1-p)^{k-1} P(Y = k)$$

$$= \sum_{k=1}^{\infty} (1-p)^{k-1} p'(1-p')^{k-1}$$

$$= p' \sum_{k=1}^{\infty} ((1-p)(1-p'))^{k-1} = \frac{p'}{1-(1-p)(1-p')}$$

$$= \frac{p'}{p' + p(1-p')} \ .$$

Sostituendo i valori noti per p e p', si trova infine $P(X \geq Y) = 3/4$.

$\triangleright$ **Esercizio 1.22**

Si lanciano una moneta e un dado non truccati. Se la moneta dà testa, si lancia il dado e si pone uguale a X il valore della faccia uscita. Se invece la moneta dà croce, si lancia il dado due volte e si pone uguale a X il minimo dei valori ottenuti nei due lanci.

(i) Trovare la densità discreta della v.a. X.

(ii) Risolvere il punto (i), nel caso in cui la moneta è truccata e la probabilità che esca testa in un lancio della moneta è p.

Soluzione

(i) La v.a X può assumere ciascuno dei valori $1, 2, 3, 4, 5, 6$. Indichiamo con T l'evento che la moneta dia testa e con C quello che esca croce. Per il teorema delle probabilità totali risulta, per $k = 1, 2, 3, 4, 5, 6$:

$$P(X = k) = P(\{X = k\} \cap T) + P(\{X = k\} \cap C) =$$

$$P(\{X = k\}|T)P(T) + P(\{X = k\}|C)P(C). \qquad (*)$$

Siccome il dado è perfetto, risulta $P(\{X = k\}|T) = 1/6$.

Per calcolare $P(\{X = k\}|C)$ procediamo così. Indichiamo con U_1 e U_2 i risultati dei due lanci indipendenti del dado, nel caso che sia uscita croce; si ha $P(U_1 = n, U_2 = m) = 1/36$, per $n, m = 1, 2, 3, 4, 5, 6$. Allora risulta, se $U \doteq \min(U_1, U_2)$:

$$P(U \geq k) = P(U_1 \geq k, U_2 \geq k) = P(U_1 \geq k) \cdot P(U_2 \geq k)$$

(per l'indipendenza delle v.a. U_1 e U_2). Questa probabilità è uguale a:

$$[1 - P(U_1 < k)][1 - P(U_2 < k)] = \left(1 - \frac{k-1}{6}\right)^2 ,$$

essendo $k - 1$ il numero dei casi favorevoli a ciascuno degli eventi $\{U_1 < k\}$ e $\{U_2 < k\}$. Analogamente $P(U > k) = (1 - \frac{k}{6})^2$. Dunque, si ha:

$$P(\{X = k\}|C) = P(U = k) = P(U \geq k) - P(U > k) =$$

$$= \left(1 - \frac{k-1}{6}\right)^2 - \left(1 - \frac{k}{6}\right)^2 = \frac{13 - 2k}{36} \ .$$

Infine, essendo $P(T) = P(C) = 1/2$, sostituendo in $(*)$ si ottiene:

$$P(X = k) = \frac{1}{6} \cdot \frac{1}{2} + \frac{13 - 2k}{36} \cdot \frac{1}{2} = \frac{19 - 2k}{72} \ .$$

(ii) Se la moneta è truccata e $P(T) = p$, riprendendo i calcoli della riga precedente, si ottiene invece:

$$P(X = k) = \frac{1}{6} \cdot p + \frac{13 - 2k}{36} \cdot (1 - p) = \frac{13 - 7p + 2k(p - 1)}{36} \ .$$

▷ Esercizio 1.29

Un'azienda ha 1000 dipendenti, dei quali $u = 400$ uomini e $d = 600$ donne. All'ufficio del personale vengono esaminate (dopo averle scelte a caso dallo schedario) $n = 100$ schede, riguardante ciascuna un dipendente. Per l' $i-$esima scheda esaminata, poniamo $X_i = 1$, se essa si riferisce ad un dipendente uomo, altrimenti poniamo $X_i = 0$. Diciamo $X = X_1 + X_2 + \ldots + X_n$ il numero delle schede che riguardano dipendenti uomini.

(i) Supponiamo che le 100 schede vengano scelte a caso dallo schedario con reinserimento (cioè dopo che una scheda è stata estratta dall'archivio, viene di nuovo reinserita); calcolare $P(X_1 = 1)$ e $P(X_2 = 1)$. Le v.a. X_1 e X_2 sono indipendenti? E in generale le v.a. X_i sono tra loro indipendenti? Qual è la legge di X? Calcolare $E(X)$ e $Var(X)$.
(ii) Supponiamo ora che le 100 schede vengano scelte a caso, ma senza reinserimento. Rispondere di nuovo ai quesiti richiesti in (i); che cosa cambia?
(iii) Con riferimento al caso (ii), indichiamo con $S \in \{1, 2, \ldots\}$ il numero minimo di estrazioni, perché si ottenga una scheda relativa ad un dipendente uomo. La v.a. S è l'istante di primo successo in una successione di prove dipendenti, in cui la probabilità del successo (estrazione scheda di un dipendente uomo) non è costante, ma dipende dalla prova stessa (schema ipergeometrico). Trovare la legge di S.

Soluzione
(i) Se la scelta delle schede avviene con reinserimento, risulta ovviamente:

$$P(X_i = 1) = \frac{u}{u+d} \doteq p, \ \ i = 1, 2, \ldots, n$$

e le v.a. X_i sono indipendenti e Bernoulliane. La v.a. $X = X_1 + X_2 + \ldots + X_n$ è binomiale di parametri n e p. Risulta quindi $E(X) = np = 100 \cdot \frac{400}{1000} = 40$ e $Var(X) = np(1-p) = 40 \cdot \frac{600}{1000} = 24$.
(ii) Se la scelta avviene senza reinserimento, X risulta essere una v.a. ipergeometrica. Si ha ancora:

$$P(X_i = 1) = p, \quad i = 1, 2, \ldots, n \qquad (*)$$

Infatti, risulta ovviamente $P(X_1 = 1) = p$; verifichiamo ora che $P(X_2 = 1)$. Si ha:

$$P(X_2 = 1) = P(X_2 = 1|X_1 = 0)P(X_1 = 0) + P(X_2 = 1|X_1 = 1)P(X_1 = 1) =$$

$$\frac{u}{u+d-1} \cdot \frac{d}{u+d} + \frac{u-1}{u+d-1} \cdot \frac{u}{u+d} = \ldots = \frac{u}{u+d} = p.$$

Ripetendo questo argomento, con un po' di pazienza si verifica che $P(X_i = 1) = p$, per ogni $i = 1, \ldots, n$ e quindi X_i risulta una v.a. Benoulliana, con $E(X_i) = p$, come nel punto (i) e $(*)$ è verificata. Le v.a. X_i non risultano però indipendenti. Infatti, ad esempio:

$$P(X_1 = 1, X_2 = 1) = P(X_2 = 1|X_1 = 1)P(X_1 = 1) =$$

$$= \frac{u-1}{u+d-1}\frac{u}{u+d} \neq \left(\frac{u}{u+d}\right)^2 = P(X_1 = 1)P(X_2 = 1).$$

Passiamo a calcolare la media e la varianza di $X = X_1 + X_2 + \ldots + X_n$ (che è ora ipergeometrica). Per la media risulta ancora $E(X) = E(X_1) + E(X_2) + \ldots + E(X_n) = np$, come nel caso (i). Per la varianza, abbiamo:

$$Var(X) = Var(X_1 + X_2 + \ldots + X_n) \neq Var(X_1) + Var(X_2) + \ldots + Var(X_n),$$

poiché le v.a. X_i sono dipendenti. Notiamo però che, siccome le v.a. X_i assumono solo valori 0 e 1, risulta:

$$E(X_1X_2) = P(X_1 = 1, X_2 = 1) = P(X_2 = 1|X_1 = 1)P(X_1 = 1) =$$

$$= \frac{u-1}{u+d-1} \cdot \frac{u}{u+d}$$

Inoltre:

$$E(X_1X_3) = P(X_1 = 1, X_3 = 1) =$$

$$= P(\{X_1 = 1, X_3 = 1\}|X_2 = 0) \cdot P(X_2 = 0)+$$

$$+P(\{X_1 = 1, X_3 = 1\}|X_2 = 1) \cdot P(X_2 = 1) =$$

$$= \frac{u}{u+d} \cdot \frac{d}{u+d-1} \cdot \frac{u-1}{u+d-2} + \frac{u}{u+d} \cdot \frac{u-1}{u+d-1} \cdot \frac{u-2}{u+d-2} =$$

$$= \frac{u(u-1)}{(u+d)(u+d-1)} \ .$$

In generale, si può mostrare che, per ogni $i < j$, risulta:

$$E(X_iX_j) = \frac{u(u-1)}{(u+d)(u+d-1)} \ .$$

Abbiamo anche:

$$Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) =$$

$$\frac{u(u-1)}{(u+d)(u+d-1)} - \left(\frac{u}{u+d}\right)^2 =$$

$$= \ldots = \frac{-ud}{(u+d)^2(u+d-1)} \ .$$

Dunque:

$$Var(X) = Var(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} Var(X_i) + 2\sum Cov(X_i, X_j),$$

dove la somma è estesa a tutti gli indici i e j ($i, j = 1, \ldots, n$), tali che $i < j$. Essi sono esattamente

$$\binom{n}{2} = \frac{n(n-1)}{2} \ .$$

Allora, siccome $Var(X_i) = \frac{u}{u+d}(1 - \frac{u}{u+d})$, si ottiene:

$$Var(X) = n \cdot \frac{u}{u+d}\left(1 - \frac{u}{u+d}\right) + n(n-1) \cdot \frac{-ud}{(u+d)^2(u+d-1)} =$$

$$= \ldots = \frac{nud}{(u+d)^2} \cdot \frac{u+d-n}{u+d-1} \ .$$

Se $u+d \to \infty$, si ottiene:

$$Var(X) \longrightarrow \frac{nud}{(u+d)^2} \ ,$$

come nel caso binomiale del caso (i).
Nel nostro caso, sostituendo i valori $u = 400$, $d = 600$ e $n = 100$, otteniamo:

$$Var(X) = \frac{100 \cdot 400 \cdot 600}{(1000)^2} \cdot \frac{900}{999} = 24 \cdot \frac{900}{999} \cong 24 \cdot 0.9009 \cong 21.621 \ .$$

Questo risultato non differisce molto dalla varianza di X trovata nel caso (i), in quanto $u + d = 1000 >> 1$.
(iii) Ragionando come nell'Esercizio 1.12, si ottiene:

$$P(S = k) = \frac{\binom{u}{0}\binom{d}{k-1}}{\binom{1000}{k-1}} \cdot \frac{u}{1000-(k-1)} =$$

$$= \frac{\binom{d}{k-1}}{\binom{1000}{k-1}} \cdot \frac{u}{1001-k} \ .$$