Problems on discrete random variables

Ignacio Cascos, Department of Statistics, Universidad Carlos III de Madrid

2018/19

1. A random experiment consists on rolling a six-face die. Assume that the probability of each face is proportional to the number shown on it. The result of the roll is denoted by random variable X.

    1. Describe the distribution of X. $P(X=k)=ck$ for $k\in\{1,2,\dots,6\}$ and some constant c. After the properties of the probability mass function, $c\geq 0$ and $\sum_{k=1}^6 P(X=k)=1$, so $c=1/21$. The distribution of X is given in the table below

k	$p_X(k)=P(X=k)$
1	1/21
2	2/21
3	3/21
4	4/21
5	5/21
6	6/21

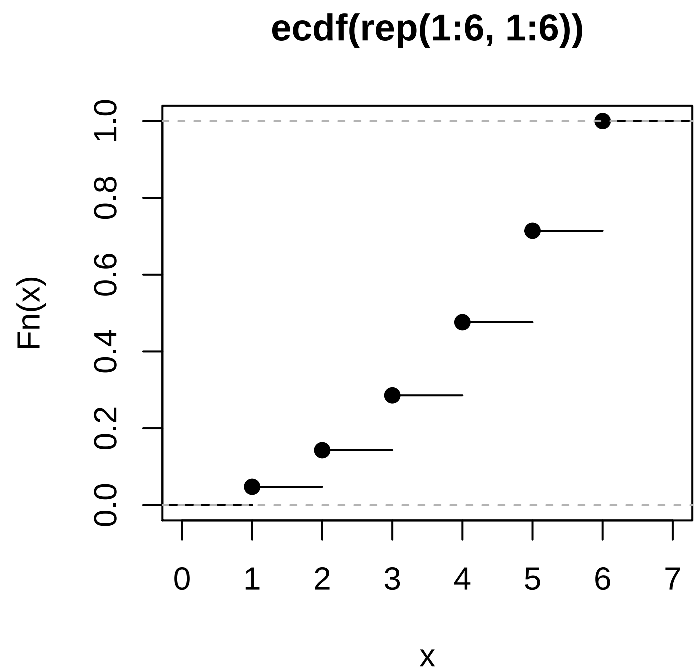
    2. Compute $\mathbb{E}[X],\mathrm{Var}[X]$, and $\sigma(X)$

    $$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^6 kp_X(k)=\sum_{k=1}^6 k\times k/21=13/3=4.33\\ \mathbb{E}[X^2] &= \sum_{k=1}^6 k^2p_X(k)=\sum_{k=1}^6 k^2\times k/21=21\\ \mathrm{Var}[X] &= \mathbb{E}[X^2]-(\mathbb{E}[X])^2=20/9=2.22\\ \sigma(X) &= \sqrt{\mathrm{Var}[X]}=\sqrt{20/9}=1.4907.\end{aligned}$$

    3. Compute the expression of the cumulative distribution function of X and represent it graphically.

    $$F_X(x)=\begin{cases} 0 & \text{if } x<1\\ 1/21 & \text{if } 1\leq x<2\\ 3/21 & \text{if } 2\leq x<3\\ 6/21 & \text{if } 3\leq x<4\\ 10/21 & \text{if } 4\leq x<5\\ 15/21 & \text{if } 5\leq x<6\\ 1 & \text{if } x\geq 6\end{cases}.$$

```r
plot(ecdf(rep(1:6,1:6)))
```



4. Compute the probability of obtaining an even number. What is the probability of obtaining an odd number? We roll the die twice, what is the probability of obtaining an even number at the first roll and an odd one at the second?

$$P(\text{'even'}) = P(X = 2) + P(X = 4) + P(X = 6) = \frac{12}{21} = \frac{4}{7}$$
$$P(\text{'odd'}) = 1 - P(\text{'even'}) = \frac{3}{7}$$
$$P(\text{'}X_1 \text{ even'} \cap \text{'}X_2 \text{ odd'}) = \frac{4}{7} \times \frac{3}{7} = \frac{12}{49}.$$

2. Two players roll two independent six-face fair dice. The result of the die of the first player is denoted by X_1, while the result of the die of the second player is X_2. The average result is denoted by A, the maximum by M, and the absolute difference by D.

$$A = \frac{X_1 + X_2}{2} \quad ; \quad M = \max\{X_1, X_2\} \quad ; \quad D = |X_1 - X_2|\,.$$

1. Determine the distribution of X_1 and the one of the random vector (X_1, X_2). Compute the mean and variance of A.

$$P(X_1 = k) = \frac{1}{6} \text{ if } k \in \{1, 2, 3, 4, 5, 6\}$$
$$P(X_1 = k_1, X_2 = k_2) = \frac{1}{36} \text{ if } k_1, k_2 \in \{1, 2, 3, 4, 5, 6\}$$
$$\mathbb{E}[A] = \mathbb{E}[(X_1 + X_2)/2] = (\mathbb{E}[X_1] + \mathbb{E}[X_2])/2 = (3.5 + 3.5)/2 = 3.5$$
$$\text{Var}[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 91/6 - (7/2)^2 = 35/12$$
$$\text{Var}[A] = \text{Var}[(X_1 + X_2)/2] = (\text{Var}[X_1] + \text{Var}[X_2])/4 = 35/24.$$

2. A player wins if her result is strictly greater than the result of her competitor. If both results are identical there is a draw. Compute the probability that player 1 wins and the probability of draw (observe that one probability is easily deduced from the other).

$$P(X_1 > X_2) = \sum_{k_1=2}^{n} \sum_{k_2=1}^{k_1} P(X_1 = k_1, X_2 = k_2) = 15/36 = 5/12$$
$$P(X_1 = X_2) = 1 - P(X_1 > X_2) - P(X_2 > X_1) = 1/6.$$

3. What is the distribution of M? Compute its mean and variance. Compare $\mathbb{E}[M]$ with $\mathbb{E}[A]$. How can you determine which of the two is greater without any computation?

$$P(M = k) = \sum_{k_1+k_2=k} P(X_1 = k_1, X_2 = k_2), \text{ if } k \in \{1, 2, 3, 4, 5, 6\}$$

k	$p_M(k) = P(M = k)$
1	1/36
2	3/36
3	5/36
4	7/36
5	9/36
6	11/36

$$\mathbb{E}[M] = \sum_{k=1}^{6} k p_M(k) = \frac{161}{36} = 4.4722$$
$$\mathbb{E}[M^2] = \sum_{k=1}^{6} k^2 p_M(k) = \frac{791}{36}$$
$$\mathrm{Var}[M] = \mathbb{E}[M^2] - \mathbb{E}[M]^2 = \frac{2555}{36^2} = 1.97145.$$

Clearly $\mathbb{E}[A] < \mathbb{E}[M]$, as it should be expected, since no realization of the average is greater than the maximum, $A \leq M$.

4. Are random variables A and M independent?

A and M are NOT independent. If M=1, then $A = 1$, while if $M = 6$, then A assumes a value in the set $\{3.5, 4, 4.5, 5, 5.5, 6\}$.

5. Compute the mean of D (observe that it is possible to write D in terms of the maximum and minimum of X_1 and X_2, or their maximum and average).

$$D = |X_1 - X_2| = \max\{X_1, X_2\} - \min\{X_1, X_2\} = 2\left(\max\{X_1, X_2\} - \frac{X_1 + X_2}{2}\right),$$
$$\mathbb{E}[D] = 2\left(\mathbb{E}[\max\{X_1, X_2\}] - \mathbb{E}\frac{X_1 + X_2}{2}]\right) = 2\left(\frac{161}{36} - \frac{7}{2}\right) = \frac{35}{18} = 1.944.$$

3. Show that if X is a discrete random variable, then

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \min_{x\in\mathbb{R}} \mathbb{E}[(X - x)^2].$$

$$\mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[(X-x+x-\mathbb{E}[X])^2] = \mathbb{E}[(X-x)^2 + (x-\mathbb{E}[X])^2 + 2(X-x)(x-\mathbb{E}[X])]$$
$$= \mathbb{E}[(X-x)^2] + (x-\mathbb{E}[X])^2 + 2(\mathbb{E}[X]-x)(x-\mathbb{E}[X]) = \mathbb{E}[(X-x)^2] - (x-\mathbb{E}[X])^2 \leq \mathbb{E}[(X-x)^2].$$

4. Show that if X and Y are two uncorrelated random variables ($\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$), then

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y].$$

$$\text{Var}[X+Y] = \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}\left[X+Y\right])^2 = \mathbb{E}\left[X^2+Y^2+2XY\right] - (\mathbb{E}[X]+\mathbb{E}[Y])^2$$
$$= \mathbb{E}\left[X^2\right] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] - (\mathbb{E}[X])^2 - (\mathbb{E}[Y]))^2 - 2\mathbb{E}[X]\mathbb{E}[Y] = \text{Var}[X] + \text{Var}[Y].$$

5. Four flatmates Alice, Bob, Charly, and Dave roll a 6-sided die every night in order to decide who washes the dishes after dinner. If the outcome is 1, Alice washes, it is its 2, Bob does, while for 3 or 4 Charly must wash the dishes and for 5 or 6 it is Dave' s turn.

   1. What is the probability that Charly washes the dishes at most twice in a week (7 days)?, and the probability that he washes exactly two days in a week?

$$X \equiv \text{'number of days Charly washes the dishes in a week'}$$

$$X \sim \text{B}(n=7, p=1/3)$$

- $P(X \leq 2)$

```
pbinom(2,size=7,prob=1/3)
```

```
## [1] 0.5706447
```

- $P(X = 2)$

```
dbinom(2,size=7,prob=1/3)
```

```
## [1] 0.3072702
```

2. How many days (dinners) must we wait on average until it is Alice's turn to wash the dishes?, and the probability that Alice washes the dishes before that time?

$$Y \equiv \text{'number of dinners until Alice washes for the first time'}$$

then r.v. $Y-1$ represents the 'number of dinners before Alice washes for the first time', and

$$Y-1 \sim \mathcal{G}(p=1/6)$$

- $\mathbb{E}[Y] = \mathbb{E}[Y-1]+1 = \frac{1-p}{p}+1 = 1/p = 6\,.$
- $P(Y<6) = P(Y-1<5) = P(Y-1 \leq 4)$

```
pgeom(4,prob=1/6)
```

```
## [1] 0.5981224
```

3. How many days (dinners) must we wait until Bob washes the dishes at least 11 times with probability 0.95?, and the (exact) probability that we have to wait longer than that time?

$$Z \equiv \text{'number of dinners until Bob washes the dishes 11 times'}$$

then r.v. $Z-11$ represents the 'number of dinners at which Bob does not wash the dishes before washing them the 11-th time', and

$$Z-11 \sim \text{NB}(r=11, p=1/6)$$

- $P(Z \le z) = 0.95$, so $P(Z - 11 \le z - 11) = 0.95$

```r
qnbinom(0.95,size=11,prob=1/6)
```

```
## [1] 88
```

```r
pnbinom(88,size=11,prob=1/6)
```

```
## [1] 0.9534517
```

- $P(Z > 99)$

```r
1-pnbinom(88,size=11,prob=1/6)
```

```
## [1] 0.04654827
```

$$V \equiv \text{'number of dinners Bob washes the dishes in 99 days'}$$

$$V \sim \mathrm{B}(n = 99, p = 1/6)$$

- $P(V \ge 11) \ge 0.95$

```r
1-pbinom(10,size=99,prob=1/6)
```

```
## [1] 0.9534517
```

- $P(V \le 10)$

```r
pbinom(10,size=99,prob=1/6)
```

```
## [1] 0.04654827
```

6. We have a lot of 100 electronic devices, out of which 20 are defective.
    1. If 20 devices are selected at random, what is the probability that four or more of them are defective?

$$X \equiv \text{'number of defective devices out of 20 (lot of 100 with 20 defectives) drawn without replacement'}$$

$$X \sim \mathrm{H}(N_1 = 20, N_2 = 80, k = 20)$$

- $P(X \ge 4) = 1 - P(X \le 3)$

```r
1-phyper(3,m=20,n=80,k=20)
```

```
## [1] 0.6083708
```

2. If 20 devices are selected at random, what is the number of non-defective devices that I can guarantee that I have with probability 0.95?

$$Y \equiv \text{'number of nondefectives out of 20 devices (lot of 100 with 20 defectives) drawn without replacement'}$$

$$Y \sim \mathrm{H}(N_1 = 80, N_2 = 20, k = 20)$$

- $P(Y \ge x) \ge 0.95$

```r
qhyper(0.95,m=80,n=20,k=20)
```

```
## [1] 18
```

```r
1-phyper(18,m=80,n=20,k=20)
```

```
## [1] 0.04984803
```

3. How many devices do we have to pick at random until we have 10 non-defective ones with probability 0.95?

   $V \equiv$ 'number of nondefectives out of k devices (lot of 100 with 20 defectives) drawn without replacement'

$$V \sim \mathrm{H}(N_1 = 80, N_2 = 20, k = k)$$

- $1 - P(Y \leq 9) = P(Y \geq 10) \geq 0.95$

```
1-phyper(9,m=80,n=20,k=15)
```

```
## [1] 0.9538525
```

7. The table below shows the number of claims reported by the owners of 15000 cars to an insurance company during a one-year period[1]. As seen in the table 11558 car owners reported no accident, while 4 car owners reported 7 accidents.

Number of car accidents	Counts
0	11558
1	2365
2	743
3	223
4	78
5	19
6	10
7	4

1. What is the average number of accidents suffered by a car in the study?

```
x=rep(0:7,c(11558,2365,743,223,78,19,10,4))
mean(x)
```

```
## [1] 0.3343333
```

2. Assume that the number of accidents suffered by a car follows a Poisson distribution and fix as λ the previous average. Obtain the probability mass function for values of r between 0 and 7 and then convert those probabilities into expected counts multiplying times 15000. Simulate 15000 observations of this distribution model and represent the counts in a barplot (use NIU, DNI, or NIE as seed).

```
dpois(0:7,lambda=mean(x))
```

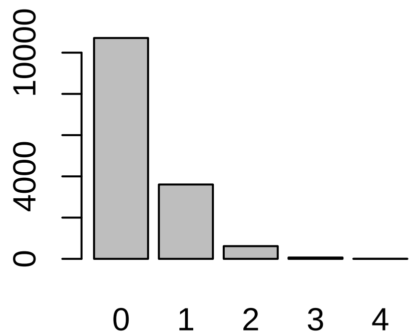
```
## [1] 7.158151e-01 2.393209e-01 4.000647e-02 4.458499e-03 3.726562e-04
## [6] 2.491828e-05 1.388502e-06 6.631749e-08
```

```
dpois(0:7,lambda=mean(x))*15000
```

```
## [1] 1.073723e+04 3.589813e+03 6.000971e+02 6.687748e+01 5.589843e+00
## [6] 3.737742e-01 2.082753e-02 9.947624e-04
```

```
set.seed(1)
barplot(table(rpois(15000,lambda=mean(x))))
```

[1]Source: M.C. Melgar (2004) AnÃ¡lisis de las Componentes de la Demanda de Seguro. AplicaciÃ³n al Seguro del AutomÃ³vil. PhD Thesis. Universidad de Sevilla.



3. The Zero Inflated Poisson (ZIP) distribution has two parameters $0 < p < 1$ and $\lambda > 0$. A ZIP(p, λ) random variable equals 0 with probability p and a Poisson $\mathcal{P}(\lambda)$ random variable with probability $1 - p$. What is the support of such a random variable? Obtain its probability mass function, mean, and variance.

Consider $Y \sim \text{ZIP}(p, \lambda)$. Clearly its support is the same to the one of a Poisson random variable, $Y(S) = \{0, 1, 2, \ldots\}$.

$$
\begin{aligned}
P(Y = 0) &= p + (1-p)e^{-\lambda} \\
P(Y = k) &= (1-p)e^{-\lambda}\frac{\lambda^k}{k!} \text{ if } k \in \{1, 2, 3, \ldots\} \\
\mathbb{E}[Y] &= p \times 0 + (1-p)\lambda = (1-p)\lambda \\
\mathbb{E}[Y^2] &= p \times 0 + (1-p)(2\lambda) = (1-p)(\lambda + \lambda^2) \\
\text{Var}[Y] &= 2(1-p)\lambda - (1-p)^2\lambda^2 = (1 + p\lambda)(1-p)\lambda
\end{aligned}
$$

4. What are the expected counts for a ZIP distribution with parameters $p = 0.6361789$ and $\lambda = 0.9189491$?

```r
p=0.6361789
(p*rep(1:0,c(1,7))+(1-p)*dpois(0:7,lambda=0.9189491))*15000
```

```
## [1] 11719.814784  2000.672834   919.258250   281.583847    64.690306
## [6]    11.889420     1.820962     0.239053
```