

Probability Revision I

Exercise I

1) $\mathbb{P}(\text{Dice} = 4) = \frac{1}{6}$

2) $\mathbb{P}(\text{Dice} \in \{4, 5, 6\}) = \frac{1}{2}$

3) We note the three dices D_1, D_2 and D_3 . We need to find

$$\mathbb{P}(D_1 + D_2 + D_3 \geq 4).$$

1st Method: $\{D_1 + D_2 + D_3 \geq 4\} = \{D_1 \geq 2\} \cup \{D_1 = 1, D_2 \geq 2\} \cup \{D_1 = 1, D_2 = 1, D_3 \geq 2\}$

with all the set disjoint.

then: $\mathbb{P}(D_1 + D_2 + D_3 \geq 4) = \mathbb{P}(D_1 \geq 2) + \mathbb{P}(D_1 = 1, D_2 \geq 2) + \mathbb{P}(D_1 = 1, D_2 = 1, D_3 \geq 2)$

$$= \frac{5}{6} + \frac{1}{6} \times \frac{5}{6} + \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6}$$

$$\boxed{\mathbb{P}(D_1 + D_2 + D_3 \geq 4) = \frac{215}{216}}$$

2nd Method: We use the complement set of $\{D_1 + D_2 + D_3 \geq 4\}$

or: $\{D_1 + D_2 + D_3 \geq 4\}^c = \{D_1 + D_2 + D_3 < 4\} = \{D_1 = 1, D_2 = 1, D_3 = 1\}$

Then: $\mathbb{P}(D_1 + D_2 + D_3 \geq 4) = 1 - \mathbb{P}(\{D_1 + D_2 + D_3\}^c) = 1 - \mathbb{P}(D_1 = 1, D_2 = 1, D_3 = 1)$

$$= 1 - \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{215}{216}$$

4) The dice being fair, $\mathbb{P}\{D=i\} = \frac{1}{6}$ for all i in $\{1, \dots, 6\}$

Then the average result is given by the expectation of D :

$$\mathbb{E}[D] = \sum_{i=1}^6 \mathbb{P}(D=i) \times i = \frac{1}{6} \sum_{i=1}^6 i = \frac{1}{6} \times \frac{6 \times 7}{2}$$

then: $\boxed{\mathbb{E}[D] = \frac{7}{2}}$

5) In this case, $\mathbb{P}(D=i) = \frac{1}{n}$ for all $i \in \mathbb{I}1, n\mathbb{I}$

then: $\mathbb{E}[D] = \sum_{i=1}^n \mathbb{P}(D=i) \times i = \sum_{i=1}^n \frac{1}{n} \times i = \frac{1}{n} \times \frac{n(n+1)}{2}$

hence: $\boxed{\mathbb{E}[D] = \frac{n+1}{2}}$

6) σ_D is calculated with $\sigma_D^2 = \mathbb{E}[D^2] - (\mathbb{E}[D])^2 = \frac{1}{n} \sum_{i=1}^n i^2 - \left(\frac{n+1}{2}\right)^2$

then $\sigma_D^2 = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 =$

then: $\boxed{\sigma_D^2 = \frac{(n+1)(n-1)}{12}}$

Exercise 2

1) We must prove that $\mathbb{P}(X_n = k) = \binom{n}{k} p^k q^{n-k}$
 with $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, for all $k \in \mathbb{I}0, n\mathbb{I}$.

It's true if $n=0$ or $n=1$ for clear reasons.

If we suppose this result true at the n^{th} step, let's prove the validity of the $(n+1)^{\text{th}}$ step.

Or to get k successes at the $(n+1)^{\text{th}}$ step we have two scenarios:

- 1) We had k successes on the n^{th} first step and a failure at the $(n+1)^{\text{th}}$ step.
- 2) We had $(k-1)$ successes on the n^{th} first step and a final success on the $(n+1)^{\text{th}}$ step:

$$\{X_{n+1} = k\} = \{X_n = k; Y_{n+1} = 0\} \cap \{X_n = k-1; Y_{n+1} = 1\}$$

there: $\mathbb{P}(X_{n+1} = k) = \mathbb{P}(X_n = k, Y_{n+1} = 0) + \mathbb{P}(X_n = k-1, Y_{n+1} = 1)$

Since all the games steps are independent from each other, we have:

$$\begin{aligned} \mathbb{P}(X_{n+1} = k) &= \mathbb{P}(X_n = k) \times q + \mathbb{P}(X_n = k-1) \times p \\ &= \binom{n}{k} p^k q^{n-k} + \binom{n}{k-1} p^{k-1} q^{n-k+1} \quad \text{with the recurrence relation} \\ &= (\binom{n}{k} + \binom{n}{k-1}) p^k q^{n-k+1} = \binom{n+1}{k} p^k q^{n+1-k} \quad \text{(Pascal's triangle)} \end{aligned}$$

2) To get the standard deviation, we must interpret X_n as the sum of n Bernoulli game (independent). Although we would have to calculate $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \times k$ or $\sum_{k=0}^n \binom{n}{k} p^k q^{n-k} k^2$.

Then, we define Y_1, \dots, Y_n independent Bernoulli game of same parameter p , such that $X_n = Y_1 + \dots + Y_n$.

Hence $\mathbb{E}[X_n] = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}(Y_i) = np$.

In addition $\mathbb{E}(X_n^2) = \sum_{i=1}^n \mathbb{E}(Y_i^2) + \sum_{\substack{i \neq j \\ i, j=1 \\ i \neq j}}^n \mathbb{E}(Y_i) \mathbb{E}(Y_j)$ if we develop X_n^2

then $\mathbb{E}(X_n^2) = np + (n^2 - n)p^2$

and $\sigma_{X_n}^2 = \mathbb{E}(X_n^2) - \mathbb{E}(X_n)^2 = np + (n^2 - n)p^2 - n^2 p^2$

and

$\sigma_{X_n} = \sqrt{np(1-p)}$

(Ex 2)

3) To give an asymptotic model for X_n , we only need to recall X_n can be expressed as the sum of n independent identically distributed random variables: $X_n = \sum_{i=1}^n Y_i$ with $E(Y_i) = p$ and $\sigma^2(Y_i) = p(1-p)$.

Or, the C.L.T (Central Limit Theorem) gives the convergence in law of $\frac{X_n - np}{\sqrt{n(p(1-p))}}$ to a standard normal gaussian $\mathcal{N}(0, 1)$.

$$\frac{X_n - np}{\sqrt{n(p(1-p))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Hence X_n has approximately the same distribution as $\mathcal{N}(np, np(1-p))$

Exercise 3

1) Noting B_1 the first bullet and B_2 the second.

$$\begin{aligned} \mathbb{P}(B_1 = \text{red}, B_2 = \text{red}) &= \mathbb{P}(B_1 = \text{red}) \times \mathbb{P}(B_2 = \text{red} | B_1 = \text{red}) \\ &= \frac{r}{r+n} \times \frac{r-1}{r+n-1} \end{aligned}$$

$$\text{hence } \mathbb{P}(B_1 = r, B_2 = r) = \frac{r(r-1)}{(r+n)(r+n-1)}$$

$$\begin{aligned} 2) \mathbb{P}(B_1 = a, B_2 = b) &= \mathbb{P}(B_1 = a) \mathbb{P}(B_2 = b | B_1 = a) \\ &= \frac{a}{r+n} \times \frac{n}{r+n-1} \end{aligned}$$

$$\text{and } \mathbb{P}(B_1 = r, B_2 = b) = \frac{rn}{(r+n)(r+n-1)}$$

(E23)

3) a) With this new setup, we must calculate $\mathbb{P}(B_2 = b)$

$$\begin{aligned}\mathbb{P}(B_2 = b) &= \mathbb{P}(B_2 = b | B_1 = r) \mathbb{P}(B_1 = r) + \mathbb{P}(B_2 = b | B_1 = b) \mathbb{P}(B_1 = b) \\ &= \frac{nr}{(n+r)(n+r+d)} + \frac{n(n+d)}{(n+r)(n+r+d)} = \frac{n}{n+r}\end{aligned}$$

Hence:

$$\boxed{\mathbb{P}(B_2 = b) = \frac{n}{n+r}}$$

b) $\mathbb{P}(B_1 = b | B_2 = b) = \frac{\mathbb{P}(B_1 = b, B_2 = b)}{\mathbb{P}(B_2 = b)}$ (Definition of the conditional probability)

then, $\mathbb{P}(B_1 = b | B_2 = b) = \frac{\mathbb{P}(B_2 = b | B_1 = b) \cdot \mathbb{P}(B_1 = b)}{\mathbb{P}(B_2 = b)}$

$$= \frac{\frac{n+d}{n+r+d} \times \frac{n}{n+r}}{\frac{n}{n+r}} \text{ with a)}$$

then by simplification:

$$\boxed{\mathbb{P}(B_1 = b | B_2 = b) = \frac{n+d}{n+r+d}}$$

Exercise 4

1) We assume $\text{Var}(X_i) < +\infty$ so that the C.L.T can apply. Then $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ (where $\mu = \mathbb{E}(X_1)$ and $\sigma^2 = \text{Var}(X_1)$) converges to $\mathcal{N}(0, 1)$ in law.

then $S_n \stackrel{L}{\sim} \mathcal{N}(n\mu, n\sigma^2)$

2) We have asymptotic confidence intervals: $[\mu - \frac{c\sigma}{\sqrt{n}}, \mu + \frac{c\sigma}{\sqrt{n}}] = I_n^c$ with c chosen in a way that $\mathbb{P}(\frac{S_n}{n} \in I_n^c)$ converges to a known quantile.

then the rate of convergence is \sqrt{n} .