Q:

Roll a fair die.

(a) What is the expected number of different faces after rolling *n* times ?

(b) What is the expected number of rolls to get all the faces from "1" to "6" ?

(c) What is the probability of getting all the faces from "1" to "6" after rolling *n* times ?

(d) Roll x times until getting all the faces from "1" to "6", what's the distribution of x?

A:

(a)

Let $X$ be number of different faces after rolling *n* times. Let $X_i = 1$ if face i appears at least once; $X_i = 0$ otherwise. Then $X = \sum_{i=1}^{6} X_i$. Hence the desired quantity is

$$E(X)$$

$$= E\left(\sum_{i=1}^{6} X_i\right)$$

$$= \sum_{i=1}^{6} E(X_i) = \sum_{i=1}^{6} P(X_i = 1) = \sum_{i=1}^{6}\left(1 - \left(\frac{5}{6}\right)^n\right) = 6\left(1 - \left(\frac{5}{6}\right)^n\right).$$

(b)

Let $Y$ be number of rolls to get all the faces from "1" to "6". Let $Y_i$ = additional number rolls until the $i$-th different face appear. Then $Y = \sum_{i=1}^{6} Y_i$.

$Y_1 = 1$. Deterministic.

$P(Y_2 = k) = \left(\frac{1}{6}\right)^{k-1} \cdot \frac{5}{6}$, i.e. $Y_2$ follows geometric distribution with parameter $p_2 = \frac{5}{6}$.

…

We can observe that for $i = 2, 3, \dots, 6$,

$P(Y_i = k) = \left(\frac{i-1}{6}\right)^{k-1} \cdot \frac{7-i}{6}$, i.e. $Y_i$ follows geometric distribution with parameter $p_2 = \frac{7-i}{6}$.

Then $E(Y_i) = \frac{1}{p_2} = \frac{6}{7-i}$ for $i = 2, 3, \dots, 6$.

Note that this formula is also true for $i = 1$. Hence the desired quantity is

$$E(Y)$$

$$= E\left(\sum_{i=1}^{6} Y_i\right)$$

$$= \sum_{i=1}^{6} E(Y_i) = \sum_{i=1}^{6} \frac{6}{7-i} = 6\sum_{i=1}^{6} \frac{1}{7-i} = 6\left(\frac{1}{6} + \frac{1}{5} + \cdots + 1\right) = 14.7.$$

(c)

Let $P_n(E)$ be the desired probability, where $E$ denotes the event that "all faces show up". Obviously, $P_n(E) = 0$ if $n < 6$.

For $n \geq 6$, $E = \bigcap_{i=1}^{6} A_i$, where $A_i$ denotes the event that face i shows up.

$P_n(E) = P_n\left(\bigcap_{i=1}^{6} A_i\right) = 1 - P_n\left(\left(\bigcap_{i=1}^{6} A_i\right)^c\right) = 1 - P_n(\bigcup_{i=1}^{6} A_i^c).$

$$P_n\left(\bigcup_{i=1}^{6} A_i^c\right) = \binom{6}{1} P_n(A_1^c) - \binom{6}{2} P_n(A_1^c A_2^c) + \binom{6}{3} P_n(A_1^c A_2^c A_3^c) - \binom{6}{4} P_n(A_1^c A_2^c A_3^c A_4^c)$$

$$+ \binom{6}{5} P_n(A_1^c A_2^c A_3^c A_4^c A_5^c)$$

$$= 6\left(\frac{5}{6}\right)^n - \binom{6}{2}\left(\frac{4}{6}\right)^n + \binom{6}{3}\left(\frac{3}{6}\right)^n - \binom{6}{4}\left(\frac{2}{6}\right)^n + \binom{6}{5}\left(\frac{1}{6}\right)^n$$

$$P_n(E) = 1 - 6\left(\frac{5}{6}\right)^n + 15\left(\frac{4}{6}\right)^n - 20\left(\frac{3}{6}\right)^n + 15\left(\frac{2}{6}\right)^n - 6\left(\frac{1}{6}\right)^n$$

```
[1] 1 0
[1] 2 0
[1] 3 0
[1] 4 0
[1] 5 0
[1]  6.0000000  0.0154321
[1]  7.0000000  0.05401235
[1]  8.0000000  0.1140261
[1]  9.0000000  0.1890432
[1] 10.0000000  0.2718121
[1] 11.0000000  0.3562064
[1] 12.0000000  0.4378157
[1] 13.0000000  0.5138582
[1] 14.0000000  0.5828453
[1] 15.0000000  0.6442127
[1] 16.0000000  0.6980044
[1] 17.0000000  0.7446325
[1] 18.0000000  0.7847071
[1] 19.0000000  0.8189231
[1] 20.0000000  0.8479875
[1] 21.0000000  0.8725775
[1] 22.0000000  0.8933165
[1] 23.0000000  0.9107646
[1] 24.0000000  0.9254152
[1] 25.0000000  0.9376979
[1] 26.0000000  0.9479827
[1] 27.0000000  0.9565864
[1] 28.0000000  0.963778
[1] 29.0000000  0.9697857
[1] 30.0000000  0.9748019
```

Specially, we can look at the case in which $n = 6$. The question is what's the probability of seeing all the six faces when the die is tossed 6 times. If we don't use the above formula, we can directly calculate

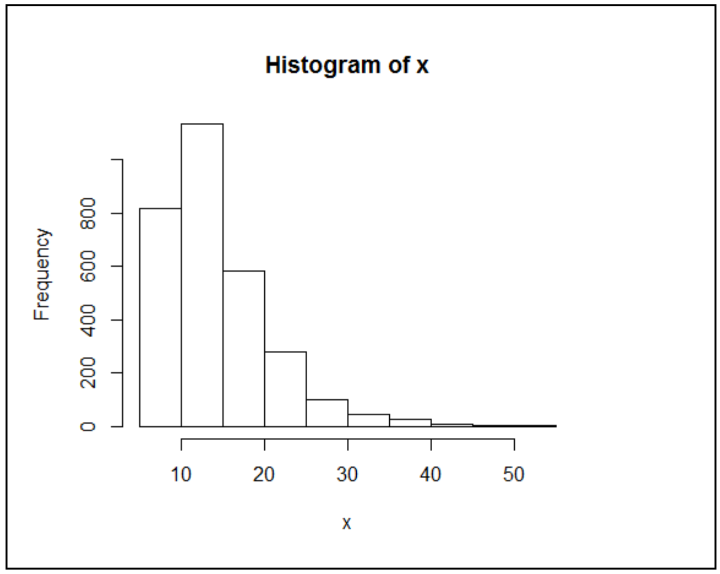
$P_6(E) = \frac{6!}{6^6} =$0.0154321.

From the numerical results, we can see that if we want to have at least 95% probability of seeing all 6 faces, we need to roll at least 27 times.

(d) Simulation method

R-code:

```
deltap = 1.0/6;
n = 3000; # Number of experiments
x = rep(0,n);
for (k in 1:n) # For the k-th experiment
{
 # Roll the fair die x times until all six faces show up
 xk = 0;
 Iface = rep(0,6);
 sum_Iface = 0;
 while (sum_Iface < 6) # While number of faces < 6
 {
  # Rolling
  U = runif(1,0,1); xk = xk + 1;
  p_sum = deltap;j = 1;
  while (p_sum < U)
  {
   p_sum = p_sum + deltap; j = j + 1;
  }
  if (Iface[j] == 0)
  {
   Iface[j] = 1;
   sum_Iface = sum_Iface + 1;
  }
 }
 x[k] = xk;
}
hist(x);
print(mean(x));
```




```
summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   6.00   10.00   13.00   14.74   18.00   51.00
```

Analytical method

Let $P_x(E)$ denote the desired probability, where $E$ denotes the event that the sixth different face shows up. Then $P_x(E) = 0$ if $x < 6$. For $x \geq 6$, we have

$$P_x(E) = \sum_{i=1}^{6} P_x(E|R_x = i)P(R_x = i) = \sum_{i=1}^{6} P_x(E|R_x = i)\frac{1}{6} = \frac{1}{6}\sum_{i=1}^{6} P_x(E|R_x = i),$$

where $R_x$ is the result of last roll.

$$P_x(E|R_x = i) = P(\text{First } x-1 \text{ rolls have all faces but } i)$$

$$= P(\text{First } x-1 \text{ rolls have all faces but } 6).$$

Hence,

$$P_x(E) = \frac{1}{6}\cdot 6 \cdot P(\text{First } x-1 \text{ rolls have all faces but } 6)$$

$$= P(\text{First } x-1 \text{ rolls have all faces but } 6)$$

$$= P_{x-1}(\cap_{i=1}^{5} A_i \cap A_6^c)$$

$$= P_{x-1}(\cap_{i=1}^{5} A_i |A_6^c)P_{x-1}(A_6^c)$$

$$= P_{x-1}(\cap_{i=1}^{5} A_i |A_6^c)\left(\frac{5}{6}\right)^{x-1},$$

where $A_i$ denotes the event that face $i$ appears and $A_i^C$ is its complementary event. Note that

$P_{x-1}\left(\bigcap_{i=1}^{5} A_i \,\middle|\, A_6^c\right)$

$= 1 - P_{x-1}\left(\left(\bigcap_{i=1}^{5} A_i\right)^c \middle| A_6^c\right)$

$= 1 - P_{x-1}\left(\bigcup_{i=1}^{5} A_i^c \,\middle|\, A_6^c\right)$

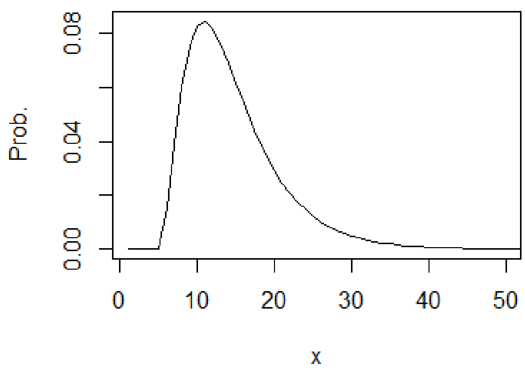
$= 1 - \binom{5}{1} P_{x-1}(A_1^c|A_6^c) + \binom{5}{2} P_{x-1}(A_1^c A_2^c|A_6^c) - \binom{5}{3} P_{x-1}(A_1^c A_2^c A_3^c|A_6^c) + \binom{5}{4} P_{x-1}(A_1^c A_2^c A_3^c A_4^c|A_6^c)$

$= 1 - 5\left(\frac{4}{5}\right)^{x-1} + 10\left(\frac{3}{5}\right)^{x-1} - 10\left(\frac{2}{5}\right)^{x-1} + 5\left(\frac{1}{5}\right)^{x-1}.$

Therefore,

$P_x(E) = \left(1 - 5\left(\frac{4}{5}\right)^{x-1} + 10\left(\frac{3}{5}\right)^{x-1} - 10\left(\frac{2}{5}\right)^{x-1} + 5\left(\frac{1}{5}\right)^{x-1}\right) \cdot \left(\frac{5}{6}\right)^{x-1}$

$= \left(\frac{5}{6}\right)^{x-1} - 5\left(\frac{4}{6}\right)^{x-1} + 10\left(\frac{3}{6}\right)^{x-1} - 10\left(\frac{2}{6}\right)^{x-1} + 5\left(\frac{1}{6}\right)^{x-1}.$




mean = 14.7

$e_i = \sum_{x=6}^{+\infty} x\left(\frac{i}{6}\right)^{x-1}$. Let $y = x - 5$, then

$e_i = \sum_{y=1}^{+\infty} (y+5)\left(\frac{i}{6}\right)^{y+4}$

$= \sum_{y=1}^{+\infty} y\left(\frac{i}{6}\right)^{y+4} + \sum_{y=1}^{+\infty} 5\left(\frac{i}{6}\right)^{y+4}$

$= \left(\frac{i}{6}\right)^5 \sum_{y=1}^{+\infty} y\left(\frac{i}{6}\right)^{y-1} + 5\left(\frac{i}{6}\right)^5 \sum_{y=1}^{+\infty} \left(\frac{i}{6}\right)^{y-1}$

$= \left(\frac{i}{6}\right)^5 \frac{6}{6-i} \sum_{y=1}^{+\infty} y\left(\frac{i}{6}\right)^{y-1} \frac{6-i}{6} + 5\left(\frac{i}{6}\right)^5 \sum_{y=1}^{+\infty} \left(\frac{i}{6}\right)^{y-1}$

$= \left(\frac{i}{6}\right)^5 \left(\frac{6}{6-i}\right)^2 + 5\left(\frac{i}{6}\right)^5 \frac{1}{1-\frac{i}{6}}$

$$= \left(\frac{i}{6}\right)^5 \left(\frac{6}{6-i}\right)^2 + 5\left(\frac{i}{6}\right)^5 \frac{6}{6-i}$$

$$= \left(\frac{i}{6}\right)^5 \frac{6}{6-i}\left(\frac{6}{6-i} + 5\right) = \left(\frac{i}{6}\right)^5 \frac{6}{6-i}\left(\frac{36-5i}{6-i}\right).$$

We can use this to calculate the mean, that is $e_5 - 5e_4 + 10e_3 - 10e_2 + 5e_1$= 14.7.