Computational-Statistics

Nina Leach

Table of contents

Preface

Part I

Probability

Pishro-Nik (2014)

Random Variables

A random experiment is a procedure where uncertain events are observed. An **outcome** is the result of such an experiment, and the collection of all possible outcomes is called the **sample space** S. For example:

- a coin toss: $S = \{heads, tails\}$
- roll a 6 sided die (D6): $S = \{1, 2, 3, 4, 5, 6\}$
- number of rainy days in June: $S = \{0, 1, ..., 30\}$

Repetitions of a random experiment is known as a **trial.** If a coin is tossed twice $S = \{(H, H), (T, H), (H, T), (T, T)\}$.

An **event** is a subset of the sample space.

The numerical results of an a random experiment are **random variables(RV)**. For the example of a coin tossed twice, numeral values can be assigned to each outcome: $(H, H) = 0, (T, H) = 1, ...$

The set of possible values for random variable X is denoted by $Range(X)$, R_X, Ω.

1. I roll a D6 10 times. Let X be the number of 1's observed. Then $R_X = \{0, 1, ..., 10\}$
2. How many attempts it takes for Manchester United (F.C.) defense to tighten up. Let Y represent the total number of games before they manage to keep a clean sheet. Then $R_Y = \{1, 2, ...\} = \mathbb{N}$ i.e. Y can be any positive integer.
3. The random variable T is the time (in hours) from now until the next earthquake occurs. This can be any non negative real number $R_T = [0, \infty)$

Part II

1 Discrete random variables

A random variable X is called discrete if:

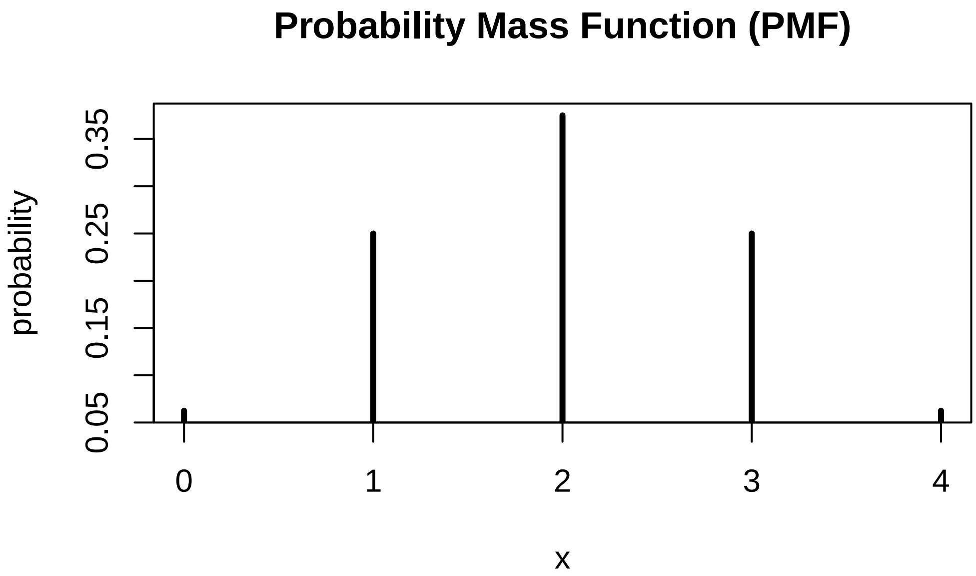
1. Ω is a finite set. E.g. $\Omega_X = \{1, 2, 3, 4\}$.
2. a countably infinite set. E.g.:
   - Natural numbers $\mathbb{N} = \{1, 2, 3, ...\}$,
   - Integers $\mathbb{Z} = \{..., -1, 0, 1, ...\}$,
   - **Rational number**, $\mathbb{Q}$ (number that can be expressed as a fraction), etc..

If Ω isn't a countable set, X is continuous.

1.0.1 Definitions explained through an example

If a coin is tossed four times:

- The sample space (S) —the list of all (2^4) possible outcomes— is $S = \{(H, H, H, H), (H, H, H, T), ...\}$.
- We define the RV, X, as the number of heads that come up in the four coin tosses. Then the range of X is $\Omega_X = \{x_1 = 0, x_2 = 1, ..., x_5 = 4\}$.
- An **event** defined as the set of outcomes $\boldsymbol{s}$ in the sample space $\boldsymbol{S}$. For example an event, **A**, could be "getting exactly 2 heads". Then **A** would include all outcomes from the sample space where there are 2 heads, e.g.: $(HHTT), (HTHT)$, etc. In general terms $A = \{s \in S | X(s) = x_k\}$
- Probabilities of events $\{X = x_k\}$ are formally shown by the probability mass function (**pmf**) of X, i.e. showing the probability of each value of Ω_X: $P(X = 0), ..., P(X = 4)$. Formally, the **pmf**: $f_X(x_k) = P(X = x_k)$.

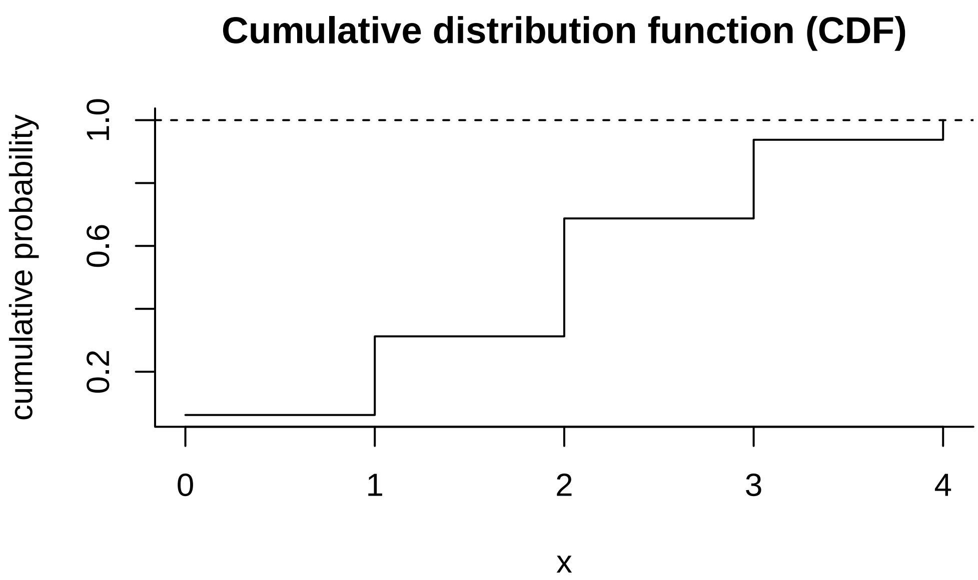


- To be a pmf, f_X must have the following properties:
  - $f_X \geq 0$ for $x \in \Omega$. i.e $P(X = x_k) < 0$
  - $\Sigma_{x \in \Omega} f_X = 1$

  From our example, for in each case $f_X \geq 0$ and the sum is 1

Number_of_Heads	Probability
0	0.0625
1	0.2500
2	0.3750
3	0.2500
4	0.0625

- The **cumulative distribution function (cdf)** $F_X(x) = P(X \leq x)$



- The cdf is a non-decreasing function for which

$$\lim_{x \to -\infty} F_X(x) = 0 \text{ and } \lim_{x \to \infty} F_X(x) = 1$$

- **expected value**

$$E(X) = \sum_{x=\Omega} x f_X(x)$$

- **variance**

$$Var(X) = E[(X - E(X))^2]$$

1.0.2 Examples

1.0.2.1 Uniform Distribution

1.0.2.2 Bernoulli Distribution

1.0.2.3 Binomial distribution

(example from Lecture 1):

Let $Y \in \{0, 1\}$ denote the outcome of a binary experiment. Where a success (y=1) has probability $p \in (0, 1)$, and failure (y=0) probability $1 - p$.

The binomial distribution $X = \sum_{i=1}^{n} Y_i$ counts the number of successes X out of n independent Bernoulli trials $Y_1, Y_2, ..., Y_n$:

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$$

$E(X) = np$

$Var(X) = np(1-p)$

In R:

```
set.seed(123)
n = 7; p = 0.4
omega = 0:n
pmf = dbinom(x = omega, size = n, prob = p)
bin.distr = data.frame(x = omega, pmf)
bin.distr
```

```
  x       pmf
1 0 0.0279936
2 1 0.1306368
3 2 0.2612736
4 3 0.2903040
5 4 0.1935360
6 5 0.0774144
7 6 0.0172032
8 7 0.0016384
```

To compute the cdf

```
# option 1
cdf1 = pbinom(q = omega, size = n, prob = p)

# option 2
cdf2 = cumsum(pmf)

all(round(cdf1, 10) == round(cdf2, 10))
```

```
[1] TRUE
```

Both options the same till 10 d.p.

```
plot(bin.distr, pch = 16, bty = "l", col = 'steelblue2',
main = 'Probability mass function')

plot(omega, cdf1, type = 's', bty = "l",
xlab = 'x', ylab = 'cumulative probability',
col = 'steelblue2', lwd = 2,
main = 'Cumulative distribution function')
abline(h = 1, lty = 2)
```

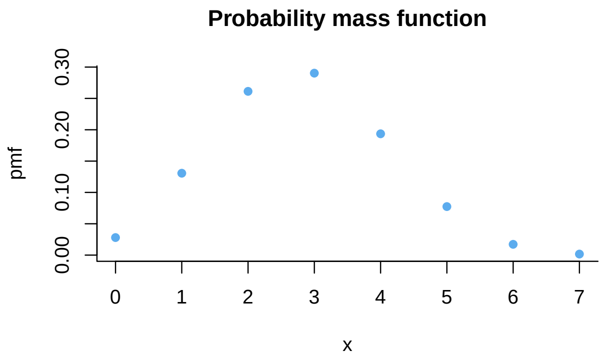


Figure 1.1: Corresponding PMF and CDF

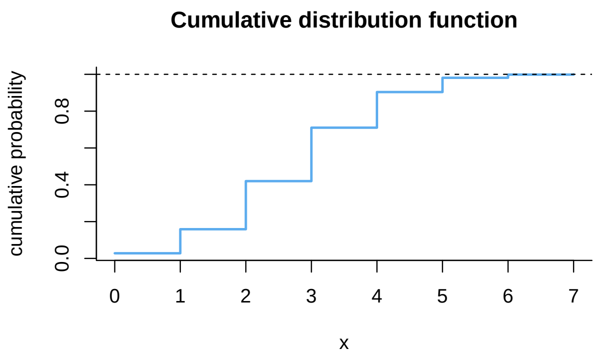


Figure 1.2: Corresponding PMF and CDF

1.0.2.4 Poisson Distribution

1.0.2.5 Geometric Distribution

1.0.2.6 Negative Binomial (Pascal) Distribution

1.0.2.7 Hypergeometric Distribution

2

3 Continuous random variables

Part III

Generating Random Numbers

4 Summary

In summary, this book has no content whatsoever.

```
[1] 2
```

A How I made this

A.1 Github

https://happygitwithr.com/

A.2 git website

https://www.codecademy.com/article/creating-a-website-on-github-pages

A.3 quarto-webr

https://quarto-webr.thecoatlessprofessor.com/

A.4 Inspiration

https://dkon1.github.io/quant_life_quarto/tutorial3.html

Pishro-Nik, Hossein. 2014. *Introduction to Probability, Statistics, and Random Processes.* Blue Bell, PA: Kappa Research, LLC.