Probability & Statistics (MA2670), Term 1 - 1st Tutorial

Dr. Dalia Chakrabarty; dalia.chakrabarty@brunel.ac.uk
Department of Mathematics, Brunel University London

- The probability distribution or cdf of r.v. X is:

$$\begin{aligned} F_X(x) &= 0 \quad x < 0, \\ &= x/4 \text{ if } \quad 0 \leq x < 1, \\ &= 1/2 \text{ if } \quad 1 \leq x < 2, \\ &= \frac{x}{12} + \frac{1}{2} \text{ if } \quad 2 \leq x < 3, \\ &= 1 \text{ if } \quad x \geq 3. \end{aligned}$$

  Using this form of the cdf, compute:

  1. $\Pr(X < 2)$.
  2. $\Pr(X = 2)$.
  3. $\Pr(1 \leq X < 3)$.
  4. $\Pr(X > 3/2)$.
  5. $\Pr(X = 5/2)$.
  6. $\Pr(2 < X \leq 7)$.
  7. Is X a discrete or continuous r.v.?

- Let a fair coin, (i.e. a coin for which probability of a head is the same as probability of a tail turning up), be tossed twice. Let the r.v. X be the number of tails that turn up in this experiment. Compute the cdf of X on the real line.

- Sales of a shop on a random day, in units of 1000GBP is the r.v. X. Cdf of X is given as:

$$\begin{aligned} F_X(x) &= 0 \quad x < 0 \\ &= (1/2)x^2 \quad 0 \leq x < 1 \\ &= k(4x - x^2) \quad 1 \leq x < 2 \\ &= 1 \quad x \geq 2. \end{aligned}$$

We have the extra information that sales are always less than 2000, i.e. $X < 2000$.

1. Find k.
2. Let the event
   $A :=$ sales tomorrow will be $\in [500, 1500]$,
   $B :=$ sales tomorrow will be > 1000,
   where all monies are in GBP.
   Then find $\Pr(A)$, $\Pr(B)$.
3. Are A and B independent events?

- Consider the function

$$\begin{aligned} g(x) &= K(2/3)^x, \text{ for } x = 1, 2, 3, \ldots \\ &= 0 \text{ elsewhere .} \end{aligned}$$

Can $g(x)$ be a probability mass function?

- *Recall the definition of cdf of a r.v. X, computed at the value a is* $\Pr(X \le a)$.

1. $\Pr(X < 2)$ is sought.
   Now $\Pr(X < 2)$ is the same as $\Pr(X \le 2)$, (i.e. $F_X(2)$), with X taking values that approach 2 arbitrarily closely, but such values are always smaller than 2. In other words, if we computed the cdf at numbers that are approaching 2 from the side of 2 that comprises values less than 2, then that computation of the cdf will yield $\Pr(X < 2)$. Formally we state:

   $$\Pr(X < 2) = \lim_{x^- \longrightarrow 2} F_X(x),$$

   the condensed notation for which is:

   $$\Pr(X < 2) = F_X(2-).$$

   Now in the given cdf, as X approaches 2 via numbers that are consistently smaller than 2, the cdf is 2/4 = 1/2. Thus

   $$\Pr(X < 2) = F_X(2-) = 1/2.$$

2. $\Pr(X = 2) = \Pr(X <= 2) - \Pr(X < 2) = F_X(2) - F_X(2-) = 2/12 + 1/2 - 1/2 = 1/6$.
3. $\Pr(1 \le X < 3) \equiv \Pr(X \in [1, 3))$. So this probability (*) should be computed at values approaching 3 via numbers that are smaller than 3, and (*) should be computed at 1, i.e. its value at 1 should be included. But if we said that $\Pr(X \in [1, 3)) = F_X(3-) - F_X(1)$, then the value of this probability at 1 is being excluded. So to include the contribution at 1, the term that is subtracted from $F_X(3-)$ needs to exclude the cdf at 1, i.e. we need to state that $\Pr(X \in [1, 3)) = F_X(3-) - F_X(1-)$ which is $= 3/12 + 1/2 - 1/4 = 1/2$.
4. $\Pr(X > 3/2) = 1 - \Pr(X \le 3/2) = 1 - F_X(3/2) = 1 - 1/2 = 1/2$.
5. $\Pr(X = 5/2) = \Pr(X \le 5/2) - \Pr(X < 5/2) = F_X(5/2) - F_X(5/2-) = 5/24 + 1/2 - (5/24 + 1/2) = 0$. This is because 5/2 is not at the edge of any branch of the cdf, i.e. the cdf displays no jumps at 5/2, so that there is no difference between the cdf at that value, and as the value is approached via numbers that are smaller than this value. Put differently, $F_X(x)$ is continuous at $x = 5/2$.
6. $\Pr(2 < X \le 7) = \Pr(X \le 7) - \Pr(X \le 2)$. The ;last term is such that the cdf is computed only at values in excess of 2, as is required for $X \in (2, 7]$. Then $\Pr(X \in (2, 7]) = F_X(7) - F_X(2) = 1 - (2/12 + 1/2) = 1/3$.

7. For a real-valued continuous r.v. X, i.e. for $X \in \mathbb{R}$ with probability density function $f_X(x)$,

   $$\Pr(X = x) = \lim_{\delta x \longrightarrow 0} \Pr(X \in [x - \delta x/2, x + \delta x/2]) = \lim_{\delta x \longrightarrow 0} f_X(x)\delta x = 0.$$

   But for a discrete r.v. X, the probability mass function is

   $$\Pr(X = x) = F_X(x) - F_X(x-) \neq 0$$

   in general.
   In this example, the r.v. X is s.t. at some x, $\Pr(X = x) \neq 0$, for example, at $x = 2$.
   But at other x, $\Pr(X = x) = 0$, as at $x = 5/2$.
   Thus this X is neither continuous, nor discrete over the whole of the real-line, but is a mixed r.v.

IMPORTANT:
Thus we see that in a cdf, the value computed at the number x may not be the same as the cdf computed as x is approached arbitrarily closely via numbers that are smaller than x, i.e. on the Real Line, as we approach x from the left of x, the computed value of the cdf need not concur with the cdf computed at x. Formally, it is possible that

$$\exists x_0 \in \mathbb{R} s.t. \lim_{x^- \longrightarrow x_0} F_X(x) \neq F_X(x_0),$$

where "$\exists$" translates to "there exists".
We say, "cdf is not necessarily left-continuous". However, if we are approaching a value via numbers that are bigger than this value, (i.e. on the Real Line, as we approach the number x from the right of x), the cdf is the same as at that value, i.e.

$$\lim_{x^+ \longrightarrow x_0} F_X(x) = F_X(x_0), \forall x_0 \in \mathbb{R}.$$

Thus one important property of the cdf is that it is right-continuous.
The cdf is also monotonically non-decreasing.
Also, it is s.t. $F_X(b) = 1$ and $F_X(a) = 0$, where $F_X(x) = 0 \forall x \in (-\infty, a)$; $x \in (b, \infty)$.

- The possible values that X can attain are 0,1,2.
  When $X = 0$, the tosses result in HH. So $\Pr(X < 0) = 0$, i.e. $F_X(0-) = 0$.
  For X to be 0, both tosses must have resulted in heads, i.e. $\Pr(X = 0) =$

$(1/2)(1/2) = 1/4$.
Thus the cdf of X must have a jump at the value 0.
When X is approaching 1 in value, through numbers that are smaller than 1, then $\Pr(X < 1)$ is still the same as $\Pr(X = 0) = 1/4$, i.e.

$$\Pr(X \in [0, 1)) = 1/4, s.t. F_X(x) = 1/4 \forall x \in [0, 1).$$

Again,

$$F_X(1) \equiv \Pr(X \leq 1) = \Pr(X = 0 \text{ or } X = 1) = \Pr(\{HH, HT, TH\}) = 3/4,$$

since there are 4 possibilities in total, namely: HH, HT, TH, TT.
Thus,

$$F_X(x) = \Pr(X \leq x), \forall x \in [1, 2).$$

Again, $\Pr(X \leq 2) = \Pr(\{HH, HT, TH, TT\})) = 1$, s.t.

$$F_X(x) \equiv \Pr(X \leq x) = 1 \forall x \geq 2.$$

Then collating over all the branches of this cdf, we get:

$$\begin{aligned} F_X(x) &= 0 \forall x < 0 \\ &= 1/4 \forall x \in [0, 1) \\ &= 3/4 \forall x \in [1, 2) \\ &= 1 \forall x \geq 2. \end{aligned}$$

- 1. We have the information that $X < 2$ in units of 1000GBP.
  Then $\Pr(X < 2) = 1$.
  Thus, $F_X(2-) = 1$, s.t. $k(4 \times 2 - 2^2) = 1$, which implies $k = 1/4$.
  2. $\Pr(A) = \Pr(0.5 \leq X \leq 1.5) = F_X(1.5) - F_X(0.5-)$
  (since we want to include the probability for X to be equal to .5).
  Thus, $\Pr(A) = (1/4)(4 \times 1.5 - 1.5^2) - (1/2)(0.5^2) = 13/16$.
  $\Pr(B) = \Pr(X > 1) = 1 - \Pr(X \leq 1) = 1 - F_X(1) = 1 - 3/4 = 1/4$..
  3. We have not seen independence yet, but you have seen this in 1st Year.
  So $\Pr(A \text{ and } B) = \Pr(0.5 \leq X \leq 1.5 \text{ and } X > 1) = \Pr(1 < X \leq 1.5) = F_X(1.5) - F_X(1)$
  (since we do not want to include the probability for X to be 1).
  So $\Pr(A, B) = 15/16 - 3/4 = 3/16$.
  Then we see that $\Pr(A)\Pr(B) \neq \Pr(A, B)$. Therefore, A and B are not independent events.

- A probability mass function (or pmf) $\Pr(X = x)$ abides by the following constraints:
  –$\Pr(X = x) = 0 \forall x \notin S$, where S is a countable set.
  –$\Pr(X = x) \geq 0 \forall x \in S$.
  –$\sum_x \Pr(X = x) = 1$, where $x \in S$.
  In this example, $g(x) = 0, \forall x \notin \mathbb{N}$ (i.e. the set of Natural numbers.
  $g(x) > 0, \forall x \in \mathbb{N}$.
  Also
  $$\sum_{x \in \mathbb{N}} K(2/3)^x = 1 \Longrightarrow K = \frac{1}{\sum_{i=1}^{\infty}(2/3)^i}.$$
  Now recalling the sum of all terms of a geometric series,
  $$\sum_{i=1}^{\infty}(2/3)^i = \frac{2/3}{1 - 2/3} = 2.$$
  Then, for $g(x)$ to sum to 1, over the set of natural numbers, we require
  $$K = 1/2$$
  Thus, for $K = 1/2$, $g(x)$ is a pmf.