Statistics 3657 : Discrete Pmf and Cdf

Some basic properties are discussed in Section 2.1

Here X denotes a generic random variable (rv). It is real valued. Note that since the integers are a subset of the real numbers this notion also includes integer valued rv's and many other discrete random variables.

If we have a sample space with a probability measure on it, then we can determine the probability distribution of a r.v. X. Notice that it does not matter which sample space we use; consider two of our coin tossing examples with different sample spaces for the game.

In general if we are only interested in calculating probabilities of events associated with a random variable then we do not need to use the sample space, but only some function needed to calculate the probabilities of these events. This notion is referred to as the distribution of a random variable, and in the next chapter the joint distribution of random variables.

For discrete random variables we use the CDF (cumulative distribution function) and PMF (probability mass function). For continuous random variables we will use the CDF and PDF (probability density function).

The distribution of a r.v. X lets us calculate values such as

$$P(X \in (a, b]) = P(a < X \leq b)$$

for intervals $(a, b]$ with $a < b$. It also allows us to calculate

$$P(X \in B)$$

for intervals $B = [a, b], B = (a, b)$ or $B = [a, b)$ or for unions of these intervals, that is the probability that X belongs to some set. One special case, useful for discrete r.v.s, is $B = \{a\} = [a, a]$.

Definition 1 *For a (real valued) random variable X, there is a corresponding cumulative distribution function (cdf), denoted F, that is related to probabilities by*

$$F(x) = P(X \leq x) = P(X \in (-\infty, x])$$

for all possible real numbers x.

Thus when we give a cdf we need to specify it as a function with domain as the real numbers, that is we need to specify the rule or function for all possible $x \in (-\infty, \infty)$.

Some properties of a cdf F are given below with an explanation or justification using the Axioms of Probability. To help with this we sometimes write an event using $\{$ and $\}$ to emphasize that we are referring to an event, which is a subset of a sample space Ω, even though we may not explicitly write out define carefully the sample space.

1. For any x, the set $\{X \le x\}$ is an event and hence $0 \le F(x) \le 1$.

2. For real numbers $x < y$ notice that $(-\infty, x] \subset (-\infty, y]$ and hence we have for events

$$\{X \le x\} \subseteq \{X \le y\}$$

   and hence $F(x) \le F(y)$. That is F is a monotone increasing function.

3. The events $\{X \le x\} \downarrow \emptyset$ as $x \downarrow -\infty$ and hence

$$\lim_{x \to -\infty} F(x) = P(\emptyset) = 0 \ .$$

   The events $\{X \le x\} \uparrow \Omega$, the sample space, as $x \uparrow \infty$ and hence

$$\lim_{x \to \infty} F(x) = P(\Omega) = 1 \ .$$

Suppose X is a rv with cdf F. Then for any $a < b$, and using properties derived from the Axioms of Probability,

$$\begin{aligned} F(b) &= P(X \le b) \\ &= P(\{X \le a\} \cup \{X \in (a, b]\}) \\ &= F(a) + P(a < X \le b) \end{aligned}$$

For the last line we are using properties of disjoint events. The student should think about where this is used.

Thus we have for any interval $(a, b]$

$$P(a < X \le b) = F(b) - F(a) \ .$$

Thus the cdf F gives the distribution of X. For any relevant sets we can calculate all probabilities needed for X. Therefore in this case we need not even worry about the sample space in any explicit sense. See the Binomial(1 , 1/2) example below.

There are two main types of rv's of interest in this course, continuous random variables and discrete random variables. There are also random variables that are neither in the sense that they are partly continuous and partly discrete. Later in the course we will consider some special types of these with some special mathematical tools.

Discrete random variables are ones that can take on, with probability 1, only values in a finite or countable set.

For example one may consider

1. the number of heads in n independent coin tosses
2. the number that comes up on a single die roll, which must be an integer number from 1 to 6
3. the sum for two die rolls, which is an integer number between 2 and 12
4. the number of independent coin tosses to obtain the first head. This one is interesting in the outcome can then be any integer from 1 onwards, that is a countable number of possible values. Let X this random variable. A variation on this game is to count the number of tails before the first head. Of course $X = Y + 1$.

   Note that both of these random variables are said to have a geometric distribution, but their distributions are different. In using the geometric distribution one should always check which of these two formulae are appropriate.

Toss 4 fair coins, and consider the two r.v.'s:

- X = number of H's (successes) in the first 3 coin tosses
- Y = number of H's (successes) in the last 2 coin tosses

It is simplest for us to obtain the joint distribution of these two discrete r.v.'s from the sample space description of this game.

elementary outcome	X	Y
0000	0	0
0001	0	1
0010	1	1
0011	1	2
0100	1	0
0101	1	1
0110	2	1
0111	2	2
1000	1	0
1001	1	1
1010	2	1
1011	2	2
1100	2	0
1101	2	1
1110	3	1
1111	3	2

In this example there are two discrete r.v.s (X, Y).

For now we just find their *marginal* distribution (this term will be discussed further in Chapter 3).

x	$P(X = x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

What is the cdf of X?

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{8} & \text{if } 0 \leq x < 1 \\ \frac{4}{8} & \text{if } 1 \leq x < 2 \\ \frac{7}{8} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

There are other standard discrete distributions, the hypergeometric and the Poisson amongst other.

In all these discrete rv cases the cdf is not continuous, and hence not differentiable everywhere. In particular it is not differentiable at each jump point. See the example below and plots Figures 2 and 4

For discrete rv's X we have a second equivalent way of writing the distribution, namely the probability mass function. It is a function, say $p : \mathcal{R} \mapsto [0,1]$ with the property that for every $x \in \mathcal{R} =$ set of real numbers then $p(x) = P(X = x)$. For a discrete r.v. $p(x) = 0$ for most values of x, and $p(x) > 0$ for only finitely or countably many values of x.

The support of a function f is

$$\{x : f(x) \neq 0\}$$

so that in the case of a pmf (or later in the case of a pdf) the support is

$$\text{support}(f) = \{x : f(x) > 0\} \ .$$

Example : X is the number of heads in a game consisting of tossing two independent coins, where each coin has probability of θ of coming up heads. Then

$$p(x) = \begin{cases} (1-\theta)^2 & \text{if} \quad x = 0 \\ 2(1-\theta)\theta & \text{if} \quad x = 1 \\ \theta^2 & \text{if} \quad x = 2 \\ 0 & \qquad \text{otherwise} \end{cases}$$

If it were a game with fair coins then $\theta = \frac{1}{2}$ and

$$p(y) = \begin{cases} \frac{1}{4} & \text{if} \quad y = 0 \text{ or } 2 \\ \frac{1}{2} & \text{if} \quad y = 1 \\ 0 & \qquad \text{otherwise} \end{cases}$$

Aside :

1. The variable or argument in the function is just a symbol to represent a real number, so it does not matter if we use x or y or any other sensible symbol as this argument.

2. For each *parameter* value θ, there is a different binomial distribution, which we write as Binom(2, θ). Of course these only give a valid distribution if $\theta \in [0,1]$. The values θ equal to 0 or 1 are usually not interesting.

3. θ is a parameter and is not the argument of the function. It only determines which particular binomial distribution we are considering.

Properties of a Probability Mass Function

The following also shows how these are consequences of the Axioms of Probability. Since we are dealing with a function, we will use below the symbol f as this function, that is f is the pmf.

1. for any x, $0 \le f(x) \le 1$

   This follows since $\{X = x\}$ is an event, and hence $0 \le f(x) = P(\{X = x\}) \le 1$.

2. $$\sum_x f(x) = 1$$

   This may also be written more carefully as

   $$\sum_{x: f(x) > 0} f(x) = 1$$

   This follows since

   $$\begin{aligned} 1 &= P(\Omega) \\ &= P(\{X \in (-\infty, \infty)\}) \\ &= P(\cup_{x: f(x)>0} \{X = x\}) \quad , \quad \text{note this is a countable union} \\ &= \sum_{x: f(x)>0} P(\{X = x\}) \\ &= \sum_{x: f(x)>0} f(x) \end{aligned}$$

There is a 1 to 1 relation between the CDF and PMF so either can be used to give the distribution of a rv X.

If we have the pmf f, then for any x we have

$$\begin{aligned} F(x) &= P(X \in (-\infty, x]) \\ &= P(\cup_{\{k: f(k) > 0 \text{ and } k \le x\}} \{X = k\}) \\ &= \sum_{k: k \le x} f(k) \end{aligned}$$

Thus the pmf allows us to calculate the CDF.

On the other hand if we have the CDF, say F, then for a x

$$F(x) = P(X \in (-\infty, x])$$

$$\begin{aligned}
&= P(\{X \in (-\infty, x)\} \cup \{X = x\}) \\
&= P(\{X \in (-\infty, x)\}) + P(\{X = x\}) \\
&= \lim_{y \uparrow x} F(y) + f(x) \\
&= F(x-) + f(x)
\end{aligned}$$

Thus the pmf evaluated x is given by

$$f(x) = F(x) - \lim_{y \uparrow x} F(y) = F(x) - F(x-) \ .$$

In terms of the graph of the cdf, this is the size of the *jump* at x. For a discrete rv the cdf is a step wise constant function with jumps are precisely the places where there are *probability masses* or possible positive probability values for the rv X.

Either the cdf or pmf can be used to give the distribution of X.

For most discrete distributions, the probability mass function (pmf) is the most common form of specifying the distribution.

Example of graphs of some 1 dimensional probability mass functions (pmf's) and cumulative distribution functions (cdf's).

Poisson Distribution

A Poisson distribution has a parameter $\lambda > 0$. This means that every $\lambda > 0$ gives rise to a different Poisson distribution. This idea of parameters is discussed more next term for statistical inference.

A Poisson distribution with parameter λ is one with pmf

$$f(k) = \begin{cases} c\frac{\lambda^k}{k!} & k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

What is the constant c?

$$1 = c\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = ce^{\lambda} \ .$$

Thus $c = e^{-\lambda}$ and the Poisson parameter λ pmf is

$$f(k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!} & k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Figures 1 and 2 shows a Poisson $\lambda = 4$ example.

The formula above is valid even if $\lambda < 0$. We then have

$$f(k) = \begin{cases} \frac{\lambda^k e^{-\lambda}}{k!} & k = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

is a function with support non-negative integers and sums to 1. However $f(1) < 0$ and $f(2) > 0$. In particular we see f does not satisfy $f(x) \geq 0$ for all x and hence is not a probability mass function.

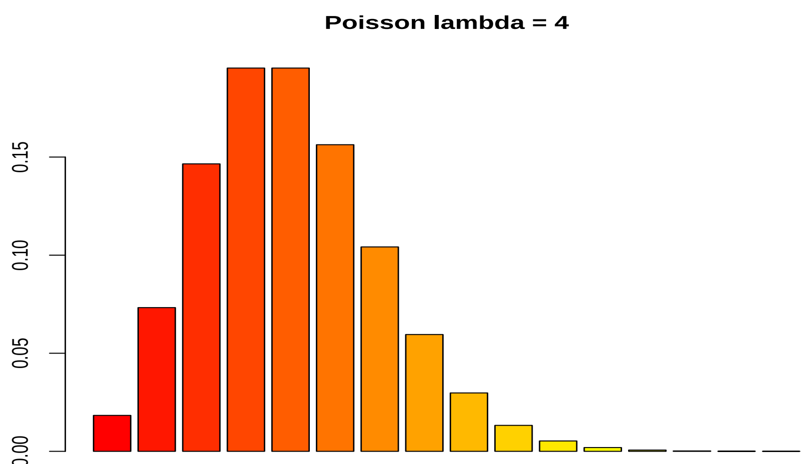

Figure 1: Poisson PMF

A Second Example

A second example is a random variable X with distribution $P(X = x) = p(x)$ where

$$p(x) = \begin{cases} c(x+6) & \text{if } x = -5, -4, \ldots, 5 \\ 0 & \text{otherwise} \end{cases}$$

Using the condition that $\sum_x p(x) = 1$ we find that $c = \frac{1}{66}$. Next check that for integers x satisfying $-5 \leq x \leq 5$ we also have $p(x) \geq 0$. Thus we have a probability mass function. Figure 3 shows the pmf and cdf. Figure 4 shows the cdf in a larger scale.

Is

$$p(x) = \begin{cases} c(x+4) & \text{if } x = -5, -4, \ldots, 5 \\ 0 & \text{otherwise} \end{cases}$$

a pmf? Using the condition that $\sum_x p(x) = 1$ we find that $c = \frac{1}{44}$. However using this value of c we do not have the condition $p(x) \geq 0$ satisfied. Actually we do not even need to calculate c. Notice $(x+4)|_{x=-5} = -1 < 0$ and $(x+4)|_{x=2} = 6 > 0$. Thus there cannot be a constant $c < 0$ or $c > 0$ that makes all these terms non-negative. If we take $c = 0$ then the sum to 1 condition is violated. Hence there is no possible value of c to make this function a probability mass function.

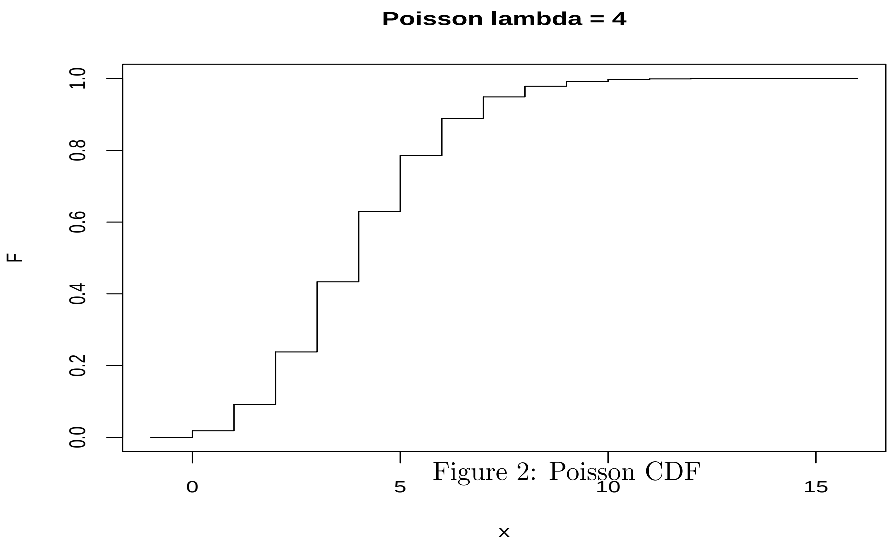

Figure 2: Poisson CDF

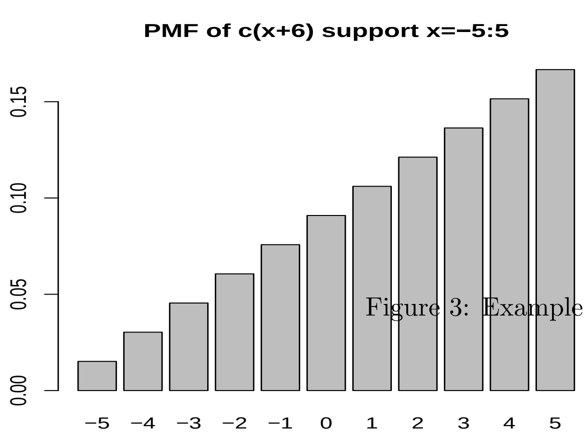

Figure 3: Example 2 PMF

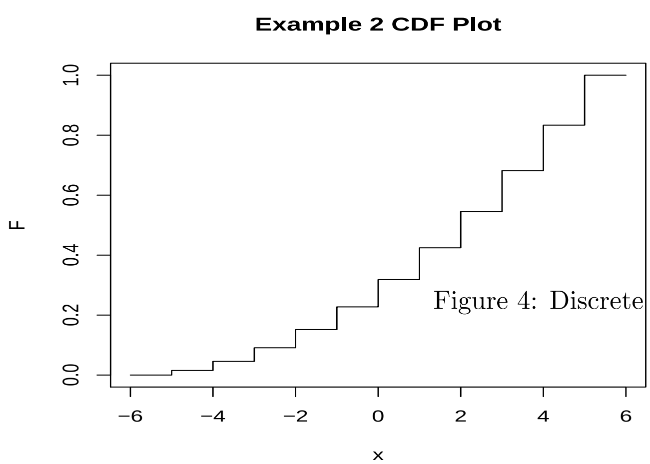

Figure 4: Discrete CDF

Binomial(2, 1/2)

One could start with a sample space

$$\Omega = \{00, 01, 10, 11\}$$

and a probability function defined to satisfy

$$P(\{\omega\}) = \frac{1}{4} \text{ , for all } \omega \in \Omega \text{ .}$$

For every possible subset A of Ω define

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

Unless we need to remind ourselves that we are dealing with sets we often write $P(\omega)$ in place of the notation $P(\{\omega\})$.

Consider the rv

$$X(\omega) = X((\omega_1, \omega_2)) = \omega_1 + \omega_2$$

It simply adds up the 1's for each elementary outcome. We can then easily calculate the pmf and then the cdf. Notice that once we have either the cdf of pmf then as far as probability calculations about X are concerned we never need to use the sample space.