Random Variables[1]

October 13, 2017

[1] HMS, 2017, v1.0

Chapter References

- Navidi: Chapter 2.4 (Not Chebyshev's inequality or the integral examples, what's important is conceptual)
- Diez: Chapter 2.4 (pure coincidence)

Random Variables

- A random variable is a number associated with the outcome of an experiment where the outcome is prone to stochastic variation.
- Throw a single six-sided die. The outcome of a throw is either 1,2,3,4,5, or 6. The random variable is the value on the face of the die and will vary on each throw.
- Associated with a random variable is a corresponding probability.
- For a given experiment, a random variable will have a range of possible values and associated probabilities.
- Random variables can be discrete or continuous

Notation: Discrete Random Variables

- Capital Roman letters, X, Y, etc indicate a random variable.
- Lower case letters, x, y, etc indicate a particular value a random variable might have.
- $P(A)$ is the probability that the event A will occur, eg number of heads.
- $P(X = x)$ is the probability that a random variable, X, takes on the value x.
- The above is also shortened to $p_X(x)$ or even to just $p(x)$ if the nature of the random variable, X, is clear to the reader.
- $p(x)$ is called the **Probability Mass Function** - PMF.

Notation: Discrete Random Variables

- Given a value x for a random variable, $p(x)$ describes its probability.

Discrete Random Variables

- For a given experiment, a random variable will have a range of possible values and associated probabilities.
- Consider the throw of a single six-sided die. The random variable, X, can have 6 values, each value is associated with a probability.

Random Variable: X	1	2	3	4	5	6
Probability $p(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

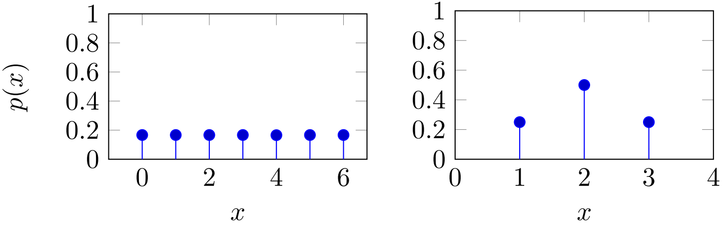
Discrete Random Variables

- Consider flipping two coins. Let the random variable, X, be the number of heads that we get during the coin flip

Outcome	HH	HT or TH	TT
Random Variable: X	2	1	0
Probability $p(x)$	1/4	1/2	1/4

Discrete Random Variables

- It is common practice to plot the probability as a function of the random variable



Such plots and tables describe a **Probability Mass Function** (PMF). The height of the line on a PMF gives the probability.

Probability Mass Function: Properties

Sum of probabilities:

$$\sum p(x) = 1$$

Mean of a PMF:

$$\mu_X = \sum x p(x)$$

Variance of a PMF:

$$\sigma_X^2 = \sum (x - \mu_X)^2 p(x)$$

Deriving a Probability Mass Function

- In a hospital, a doctor goes from bed to bed deciding which patient will be operated on that day.
- Let q = probability that a patient will be operated on.
- Let X be the patient number (ie the random variable)
- Let $p(x)$ be the probability that patient x is the first patient to be operated on. Let us derive the PMF for this situation.

$p(1) = q$

$p(2) = \overline{p(1)}$ AND $p(2) = (1-q)q$

$p(3) = \overline{p(1)}$ AND $\overline{p(2)}$ AND $p(3) = (1-q)(1-q)q = (1-q)^2 q$

Deriving a Probability Mass Function

- For the kth patient:
- $p(k) = (1-q)^{k-1}q$
- This is called the geometric distribution

Examples of Common PMFs

- Binomial Distribution
- Geometric Distribution
- Possion Distribution

Continuous Random Variables

- Because an infinite number of points lie on the x axis real line, the probability of observing any particular event is zero.
- When considering continuous random variables, we always consider the probability within a given range.

Continuous Random Variables

- $P(X \leq x)$ is the probability that a random variable, X, will be less than or equal to x.
- As x increases, $P(X \leq x)$ tends to one because we're covering the entire range.
- $P(X \leq x)$ is called the **cumulative distribution function, CDF.**
- The shorthand $F_X(x)$ is often used to depict this function.

Probability Density Function

- Define a function $f_X(x)$ that when integrated yields the CDF, $F_X(x)$.
- This means that the area under $f_X(x)$ describes the probability.
- $f_X(x)$ is often shortened to $f(x)$ and is called the **probability density function**, PDF.

Probability Density Function

Properties

$$\int_{-\infty}^{\infty} f(x)dx = 1 \qquad f(x) \geq 0 \text{ for all } x$$

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

- Mean of a PDF:

$$\mu_X = \int_{-\infty}^{\infty} x f(x)dx$$

- Variance of a PDF:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx$$

Notation: Continuous Random Variables

- Capital Roman letters, X, Y, etc indicate a random variable.
- Lower case letters, x, y, etc indicate a particular value a random variable has.
- $P(x_1 \leq X \leq x_2)$ is the probability that a random variable has a value between x_1 and x_2.
- The above is also shortened to $f_X(x)$ or even to just $f(x)$ if the nature of the random variable, X, is clear to the reader.
- $f(x)$ is often called a **Probability Density Function** - PDF.

Continuous Random Variables

- Consider a random variable: X = Ages of children from 1 to 6 years old.
- Assuming that age is equally distributed across the range.

Random Variable: X	1 to 2	2 to 3	3 to 4	4 to 5	5 to 6
Probability $P(X)$	0.2	0.2	0.2	0.2	0.2

The table describes a uniform probability distribution.

Examples of Common PDFs

- Normal Distribution
- Chi-Square
- F Distribution
- t-Distribution
- Exponential Distribution

Exercise: Part 1

- Let the random variable X be the duration of telephone calls in minutes in a busy call center.
- Is X continuous or discrete?

Exercise: Part 2

- The duration of the telephone calls is given by the PDF:

$$f(x) = 4e^{-4t}t$$

  where t is time in minutes.

- What is the probability that a call will last **less** than or equal to 30 seconds?

- Draw a diagram of the PDF

Exercise: Part 3

- The strategy is to find the area under the PDF from 0 to 30 seconds (0.5 minutes):

$$
\begin{aligned}
P(T \leq 0.5) &= \int_0^{0.5} 4e^{-4t}t \\
&= \left[-e^{-4t}\right]_0^{0.5} \\
&= 1 - e^{-2} \\
&= 0.86
\end{aligned}
$$

In software it is often easier to use the CDF function, eg in Python:
`scipy.stats.expon.cdf (0.5, scale=0.25)`