C&EE 110:
Introduction to Statistics and Probability

Einar Balan

Contents

1 Sets and Probablity Theory

1.1 Probabilistic Sets

A random event, E, has more than 1 possible outcome in the sample space S. S is the collection of all possible event outcomes. We know that E $\subset$ S.

Ex) Number in dice roll

$$S = \{1, 2, 3, 4, 5, 6\}$$
$$E_{odd} = \{1, 3, 5\}$$
$$E_{>3} = \{4, 5, 6\}$$

Operations

We can apply several operations to our sets.

1. Union, denoted $E_1 \cup E_2$
2. Intersection, denoted $E_1 \cap E_2$ or $E_1 E_2$

Consider E_{odd} and $E_{>3}$ above.

$$E_{odd} \cup E_{>3} = \{1, 3, 4, 5, 6\}$$
$$E_{odd} \cap E_{>3} = \{5\}$$

These operations are commutative, associate, and distributive. Intersection has precedence over union.

Special Events

- S is the event the spans the entire sample space
- $\varnothing$ is the null event, it has no outcomes
- if E_1 and E_2 are mutually exclusive, $E_1 E_2 = \varnothing$
- if E_1 and E_2 are collectively exhaustive, $E_1 \cup E_2 = S$
- $\overline{E_1} = S - E_1$, the complement[1] of E_1

[1]Demorgan's Laws hold

Frequentist Probability (Natural Variation)

The probability of occurrence of E is the relative frequency of observations of E ina large number of repeated experiments. Put more formally below,

$$P(E) = \lim_{N\to\infty} \frac{n}{N}, \text{ where n = occurences of E in N observations in S}$$

Bayesian Probability (Incomplete Knowledge)

The probability of an event E represents analysts' degree of belief that E will occur.

Frequentist Probability	Bayesian Probability
probability of expecting a ground shaking intensity of 1g in next 100 years	probability of finding water on new planet
max wind speed in a year	probability that a buidling will collapse under ground shaking intensity of 1g
live load on a building	election results
*based on previous observations	*not based on previous observations
*cannot be reduced through more measurement	*can be reduced if more observations/measurements applied

1.2 Axioms

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. $P(A \cup B) = P(A) + P(B)$, s.t. $AB = \varnothing$

* these axioms are consistent with Frequentist probability

We can derive several rules from these axioms.

1. $P(\overline{E}) = 1 - P(E)$
2. $P(\varnothing) = 0$
3. $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1E_2)$
   - if E_1 and E_2 are mutually exclusive, then we double count their intersection when using the 3rd axiom; subtracting it leads to the correct value
   - what if we have > 2 events? Inclusion/Exclusion rule
   - $P(E_1 \cup E_2 \cup ... \cup E_n) =$

$$\sum_{i=1}^{n} P(E_i) - \sum_{i=1}^{n}\sum_{j=1}^{i-1} P(E_iE_j) + \sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1} P(E_iE_jE_k) + ... + (-1)^{n-1}P(E_1E_2...E_n)$$

Conditional Probability

We may want to determine the probability of an event given another event is guaranteed to occur. This is denoted $P(E_1|E_2)$, which is read as E_1 given E_2. It essentially redefines the sample space to be E_2.

$$P(E_1|E_2) = \begin{cases} \frac{P(E_1E_2)}{P(E_2)} & P(E_2) > 0 \\ 0 & P(E_2) = 0 \end{cases} \tag{1.1}$$

From this equation, it follows that

$$P(E_1E_2) = P(E_1|E_2)P(E_2)$$

This holds in general for n events.

$$P(E_1E_2E_3) = P(E_1|E_2E_3)P(E_2E_3) = P(E_1|E_2E_3)P(E_2|E_3)P(E_3)$$

Ex) Applying conditions to operations

$$P(E_1 \cup E_2|E_3) = P(E_1|E_3) + P(E_2|E_3) - P(E_1E_2|E_3)$$

$$P(E_1E_2|E_3) = P(E_1|E_2|E_3)P(E_2|E_3), \text{ which follows from 1.1}$$

Independence

Two events are indpendent iff $P(E_1|E_2) = P(E_1)$

We have mutual independence if $P(E_1E_2 \dots E_n) = P(E_1)P(E_2)\dots P(E_n)$.

Theorem of Total Probability

Consider an event A and a set of of mutually exclusive and collectively exhaustive events $E_1, E_2, \dots, E_3$.

$$P(A) = \sum_{i=1}^{n} P(A|E_i)P(E_i) \tag{1.2}$$

Bayes' Rule

Consider an event A and a set of of mutually exclusive and collectively exhaustive events $E_1, E_2, \dots, E_3$ in S.

$$P(AE_j) = P(E_j|A)P(A) = P(A|E_j)P(E_j)$$

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{P(A)}$$

$$P(E_j|A) = \frac{P(A|E_j)P(E_j)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)}$$

where equation 1.2 is used to subtitute P(A)

2 | Random Variables

A **random variable** is a variable whose specific value cannot be predicted with certaintiy before an experiement. They take on a numerical value for each possible event in the sample space.

Ex) Random variables are easy to define. For example,

- X = magnitutde of a future earthquake
- Y = yield stress of a material
- Z = peak wind pressure during a given year

For a random variable X, its outcomes are denoted $x_1, x_2, \ldots, x_n$. For an outcome x_i, we denote the probability of that outcome as $P(X = x_i)$.

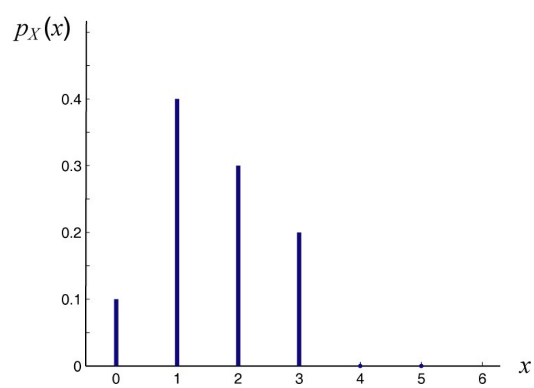
2.1 Discrete Random Variables

A random variable is called **discrete** if the number of outcomes is countable. For example, for X = the number of cars on a bridge at a certain time, X is discrete.

Distributions of discrete random variables can be quantified in 2 ways.

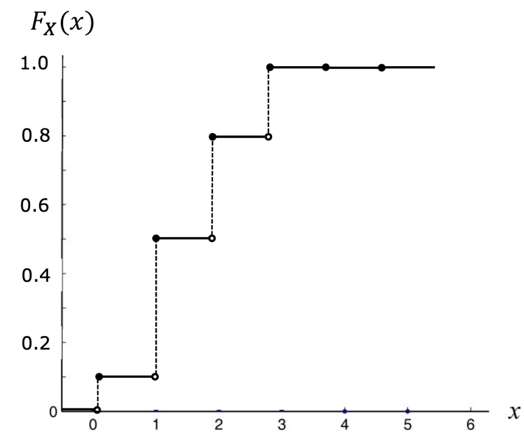
1. Probability Mass Function

$$p_X(x_i) = P(X = x_i)$$



2. Cumulative Distribution Function

$$F_X(x_i) = P(X \leq x_i)$$



Intuitively, adding up $p_X(x_i)$ for all i is equal to $F_x(a)$.

$$F_X(a) = \sum_{\text{all } x_i \leq a} p_X(x_i)$$

Rules of Discrete Random Variables

- $0 \leq p_X(x_i) \leq 1$
- $\sum_{\text{all } x_i} p_X(x_i) = 1$
- $F_X(-\infty) = 0$
- $F_X(+\infty) = 1$
- $F_X(b) \geq F_X(a)$ if $b \geq a$

All of these rules are fairly intuitive. For example, the probability of any event must be between 0 and 1. Additionally, the sum of all events in a sample space must be 1.

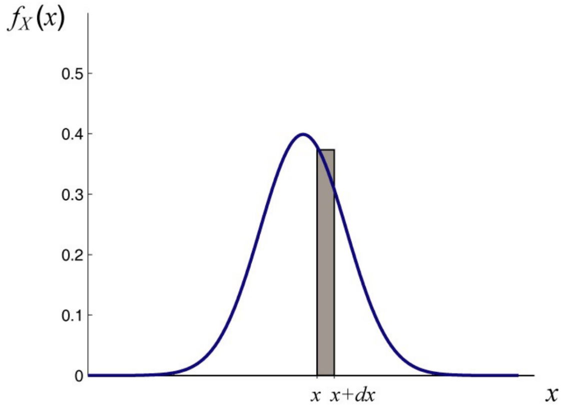
2.2 Continuous Random Variables

Random variables are said to be **continuous** if they can take on any real value. As a result, there are ∞ possible values for a random variable X. It follows that

$$P(X = x_i) = \frac{1}{\infty} = 0, \text{ for all } i$$

We can describe the distribution of continuous random variables in 2 ways.

1. Probability Density Function
$$f_X(x_i)dx = P(x_i < X < x_i + dx)$$

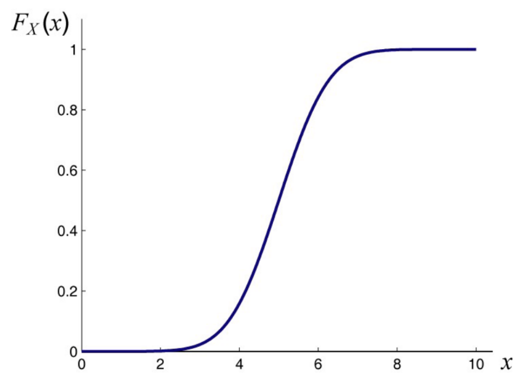


We know that occurences in different intervals are mutually exclusive, so it follows that

$$P(a < X \leq b) = \int_a^b f_X(x)dx$$

2. Cumulative Distribuion Function

$$F_X(x_i) = P(X \leq x_i)$$



Additionally,

$$F_X(x_i) = P(X \leq x_i) = \int_{-\infty}^{x_i} f_X(u)du$$

and it follows from the Fundamental Theorem of Calculus that

$$f_X(x) = \frac{d}{dx}F_X(x)$$

Rules of Continuous Random Variables

- $f_X(x) \geq 0$
- $\int_{-\infty}^{+\infty} f_X(x)dx = 1 = S$
- $F_X(-\infty) = 0$
- $F_X(\infty) = 1$
- $F_X(b) \geq F_X(a)$ if $b \geq a$

Some of these rules hold in both the discrete and the continuous case.

2.3 Joint Random Variables

Sometime it is useful to consider the probabilistic relationship between two random variables. This is called the joint probability. Just as before, we can use 2 methods to describe the joint distribution of continuous random variables.

1. Joint Probability Density Function

$$f_{X,Y}(x,y)dxdy = P(x < X \leq x + dx \cap y < Y \leq y + dy)$$

Similarly to single PDF's, we can find the probability of X and Y within a certain region as follows.

$$P(a < X \leq b \cap c < Y \leq d) = \int_a^b \int_c^d f_{X,Y}(u,v)dudv$$

Two conditions must hold for joint PDF's

(a) $f_{X,Y}(x,y) \geq 0$

(b) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(u,v)dudv$

2. Joint Cumulative Distribution Function

$$F_{X,Y}(x,y) = P(X \leq x \cap Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)dudv$$

And just as in the single variable case, the Fundamental Theoreom of Calculus can be applied to obtain

$$f_{X,Y}(x,y) = \frac{\delta^2}{\delta x \delta y} F_{X,Y}(x,y)$$

Marginal Distributions

Given the joint distribution of X and Y, we can obtain the distributions of X alone or Y alone. This is called the marginal distribution.

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

This follows from the fact that the the distribution of y from $-\infty$ to $+\infty$ is 1 and the intersection of any value with 1 is the original value.

The same style of manipulation can be performed in order to obtain the marginal CDF.

$$F_X(x) = F_X(x,\infty)$$

Conditional Probability Distributions

Conditional probability distributions are used to determine the probability of a variable given the value of another variable. Put another way, $P(X|Y)$.

For continuous random variables, the conditional PDF is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

The cumulative PDF is

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(u|y)du$$

Independence

Similar independence rules hold for random variables as in the case of events.

Specifically, X and Y are said to be independent if

$$f_{X|Y}(x|y) = f_X(x)\ \forall y$$

The following are equivalent to the above statement

- $f_{Y|X}(y|x) = f_Y(y)$
- $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
- $F_{X|Y}(x|y) = F_X(x)$
- $F_{Y|X}(y|x) = F_Y(y)$
- $F_{X,Y}(x,y) = F_X(x)F_Y(y)$

We often assume independence in order to simplify calculations.

Joint Discrete Random Variables

The rules for joint distributions also apply to discrete random variables.

Specifically,

$$p_{X,Y} = P(X = x \cap Y = y)$$

$$p_{X|Y} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

$$p_X(x) = \sum_{\text{all } y_i} p_{X,Y}(x, y_i)$$

$$p_{X|Y}(x|y) = p_X(X) \text{ given X and Y are independent}$$