STAT 330 - Mathematical Statistics

Cameron Roopnarine

Last updated: September 13, 2020

Contents

Chapter 2

Random Variable

LECTURE 1 | 2020-09-09

Review of:

- Probability
- Random variables (discrete and continuous)
- Expectation and variance
- Moment generating function

2.1 Probability Model

DEFINITION 2.1.1: Probability model

A **probability model** is used for a random experiment, which consists of three components:
(I) Sample space
(II) Event
(III) Probability function

DEFINITION 2.1.2: Sample space

A **sample space** S is a set of all the distinct outcomes for a random experiment, with the property that in a single trial, one and only one of these outcomes occurs.

EXAMPLE 2.1.3

Toss a coin twice. This is a random experiment because we do not know the outcome before we toss the coin twice.

- $S = \{(H,H),(H,T),(T,H),(T,T)\}$

Define A: First toss is an H.
Clearly, $A = \{(H,H),(H,T)\} \subseteq S$, so A is an event.

DEFINITION 2.1.4: † Sigma algebra

A collection of subsets of a set S is called **sigma algebra**, denoted by β, if it satisfies the following properties:

(I) $\varnothing \in \beta$

(II) If $A \in \beta$, then $\bar{A} \in \beta$

(III) If $A_1, A_2, \ldots \in \beta$, then $\bigcup_{i=1}^{\infty} A_i \in \beta$

DEFINITION 2.1.5: Probability set function

Let β be a sigma algebra associated with the sample space S. A **probability set function** is a function P with domain β that satisfies the following axioms:

(I) $P(A) \geqslant 0$ for all $A \in \beta$

(II) $P(S) = 1$

(III) *Additivity property*: If $A_1, A_2, A_3, \ldots \in \beta$ are pairwise mutually exclusive events; that is, $A_i \cap A_j = \varnothing$ for all $i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

EXAMPLE 2.1.6

Toss a coin twice, given one event A,

$$P(A) = \frac{\text{\# of outcomes in } A}{4}$$

since $|S| = 4$. P satisfies the three properties, therefore P is a probability function.

PROPOSITION 2.1.7: Additional Properties of the Probability Set Function

Let β be a sigma algebra associated with the sample space S and let P be a probability set function with domain β. If $A, B \in \beta$, then:

(1) $P(\varnothing) = 0$

(2) If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$

(3) $P(\bar{A}) = 1 - P(A)$

(4) If $A \subset B$, then $P(A) \leqslant P(B)$

Note for (4), $A \subset B$ means $a \in A$ implies $a \in B$.

Proof of: 2.1.7

Proof of (1): Let $A_1 = S$ and $A_i = \varnothing$ for $i = 2, 3, \ldots$. Since $\bigcup_{i=1}^{\infty} A_i = S$, then by (III) it follows that

$$P(S) = P(S) + \sum_{i=2}^{\infty} P(\varnothing)$$

and by (II) we have

$$1 = 1 + \sum_{i=2}^{\infty} P(\varnothing)$$

By (I) the right side is a series of non-negative numbers which must converge to the left side which is 1 which is finite which results in a contradiction unless $P(\varnothing) = 0$ as required.

Proof of (2): Let $A_1 = A$, $A_2 = B$, and $A_i = \varnothing$ for $i = 3, 4, \ldots$. Since $\bigcup_{i=1}^{\infty} A_i = A \cup B$, then by (III)

$$P(A \cup B) = P(A) + P(B) + \sum_{i=3}^{\infty} P(\varnothing)$$

and since $P(\varnothing) = 0$ by the result of (1) it follows that

$$P(A \cup B) = P(A) + P(B)$$

Proof of (3): Since $S = A \cup \bar{A}$ and $A \cap \bar{A} = \varnothing$ then by (II) and by (2) it follows that

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$$

as required.
Proof of (4): Since

$$B = (A \cap B) \cup (\bar{A} \cap B) = A \cup (\bar{A} \cap B)$$

and $A \cap (\bar{A} \cap B) = \varnothing$ then by (2)

$$P(B) = P(A) + P(\bar{A} \cap B)$$

But by (I), $P(\bar{A} \cap B) \geqslant 0$, so the result now follows.

EXERCISE 2.1.8

Let β be a sigma algebra associated with the sample space S and let P be a probability set function with domain β. If $A, B \in \beta$ then prove the following:

1. $0 \leqslant P(A) \leqslant 1$
2. $P(A \cap \bar{B}) = P(A) - P(A \cap B)$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1. $P(A) \geqslant 0$ follows from (I). From (3) we have $P(\bar{A}) = 1 - P(A)$. But from (I) $P(\bar{A}) \geqslant 0$ and therefore $P(A) \leqslant 1$.
2. Since $A = (A \cap B) \cup (A \cap \bar{B})$ and $(A \cap B) \cap (A \cap \bar{B}) = \varnothing$, then by (2)

   $$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

   as required.
3. $P(A \cup B) = (A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B)$. By the previous result,

   $$P(A \cap \bar{B}) = P(A) - P(A \cap B) \text{ and } P(\bar{A} \cap B) = P(B) - P(A \cap B)$$

   Therefore,

   $$\begin{aligned} P(A \cup B) &= (P(A) - P(A \cap B)) + P(A \cap B) + (P(B) - P(A \cap B)) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

   as required.

DEFINITION 2.1.9: Conditional probability

Let β be a sigma algebra associated with the sample space S and suppose $A, B \in \beta$ with $P(B) > 0$. Then the **conditional probability** of A given that B has occurred is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

DEFINITION 2.1.10: Independent events

Let β be a sigma algebra associated with the sample space S and suppose $A, B \in \beta$. A and B are **independent events** if

$$P(A \cap B) = P(A)P(B)$$

Clearly, $P(A \mid B) = P(A)$ if A and B are independent since

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

EXAMPLE 2.1.11

Toss a coin twice.

- A: First toss is H
- B: Second toss is T

$$P(A) = \frac{\text{\# of outcomes in } A}{4} = \frac{2}{4}$$

also

$$P(B) = \frac{2}{4}$$

$$P(A \cap B) = \frac{1}{4} = P(A)P(B)$$

therefore A and B are independent.

2.2 Random Variable

DEFINITION 2.2.1: Random variable

A **random variable** X is a function from a sample space S to the real numbers $\mathbb{R}$; that is,

$$X : S \to \mathbb{R}$$

satisfies for any given $x \in \mathbb{R}$ $\{X \leqslant x\}$ is an event.

$$\{X \leqslant x\} = \{\omega \in S : X(\omega) \leqslant x\} \subseteq S$$

EXAMPLE 2.2.2

Toss a coin twice. X: # of H in two tosses
Possible values of X: $0, 1, 2$. Given $x \in \mathbb{R}$.

$$\{X \leqslant x\}$$

- $x < 0$ then $\{X \leqslant x\} = \varnothing$
- $0 \leqslant x < 1$ then

then

$$\{X \leqslant x\} = \{X = 0\} = \{(T, T)\} \subseteq S$$

therefore X is a random variable.

DEFINITION 2.2.3: Cumulative distribution function

The **cumulative distribution function** (c.d.f.) of a random variable X is defined by

$$F(x) = P(X \leqslant x)$$

for all $x \in \mathbb{R}$. Note that the c.d.f. is defined for all $\mathbb{R}$

DEFINITION 2.2.4: Properties of the cumulative distribution function

(1) F is a non-decreasing function; that is, if $x_1 \leqslant x_2$, then $F(x_1) \leqslant F(x_2)$.
By looking at:
- $\{X \leqslant x_1\} \subseteq \{X \leqslant x_2\}$ if $x_1 \leqslant x_2$.

(2) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.
By looking at:
- $x \to \infty$: $\{X \leqslant x\} \to S$
- $x \to -\infty$: $\{X \leqslant x\} \to \varnothing$

(3) $F(x)$ is a right continuous function; that is, for any $a \in \mathbb{R}$,

$$\lim_{x\to a^+} F(a) = F(a)$$

(4) For all $a < b$

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

(5) For all b

$$P(X = b) = P(\text{jump at } b) = \lim_{t\to b^+} F(t) - \lim_{t\to b^-} F(t) = F(b) - \lim_{t\to b^-} F(t)$$

Lecture 2 | 2020-09-09

2.3 Discrete Random Variables

DEFINITION 2.3.1: Discrete random variable

If a random variable X can only take finite or countable values, X is a **discrete random variable**.

In this case, $F(x)$ is a right-continuous step function.

REMARK 2.3.2

When we say **countable**, we mean something you can enumerate such as $\mathbb{Z}$ or $\mathbb{N}^+$.

DEFINITION 2.3.3: Probability density function

If X is a discrete random variable, then the **probability density function** (p.d.f.) of X is given by

$$f(x) = \begin{cases} P(X = x) = F(x) - \lim_{\varepsilon\to 0^+} F(x - \varepsilon) & \text{if } X \text{ can take value } x \\ 0 & \text{if } X \text{ cannot take value } x \end{cases}$$

DEFINITION 2.3.4: Support set

The set $A = \{x : f(x) > 0\}$ is called the **support set** of X. These are all the positive values X can take.

PROPOSITION 2.3.5: Properties of the Probability Function

(1) $f(x) \geqslant 0$ *for* $x \in \mathbb{R}$
(2) $\sum\limits_{x \in A} f(x) = 1$

Review some commonly used discrete random variables:

- **Bernoulli.** $X \sim$ Bernoulli(p) where X can only take two possible values 0 (failure) or 1 (success). Let p be the probability of a success for a single trial. So,

$$P(X = 1) = p \text{ and } P(X = 0) = 1 - p$$

  Therefore,

$$f(x) = P(X = x) = p^x (1-p)^{1-x}$$

  Example: Toss a coin twice. Let X be the number of heads. Then $X \sim$ Bernoulli(p)

- **Binomial.** $X \sim \text{Binomial}(n, p)$. Suppose we have **Bernoulli Trials**:
  - We run n trials
  - Each trial is independent of each other
  - Each trial has two possible outcomes: 0 (failure), 1 (success)

$$P(X = 1) = p$$

  Let X be the number of success across these n trials and p be the success probability for a single trial.

$$X = \sum_{i=1}^{n} X_i$$

  X_i is the outcome of the ith trial.

$$P(X_i = 1) = p$$

  where $X_i \sim$ Bernoulli(p). Therefore,

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- **Geometric.** $X \sim \text{Geometric}(p)$. Let X be the number of failures before the first success. X can take values $0, 1, 2, \ldots$.

$$f(x) = P(X = x) = (1-p)^x p$$

  Example. $X =$ number of tails before you get the first head.

- **Negative Binomial.** $X \sim \text{NB}(r, p)$. Let X be the number of failures before you get r success. X can take values $0, 1, 2, \ldots$

$$f(x) = P(X = x) = \binom{x + r - 1}{x} (1-p)^x p^{r-1} p$$

  Example. $X =$ number of tails before you get the rth head.

- **Poisson.** $X \sim \text{Poisson}(\mu)$ where $X = 0, 1, \ldots$.

$$f(x) = P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$$

  where $x = 0, 1, 2, \ldots$.

EXERCISE 2.3.6

Verify all that all the probability models above are indeed probability functions using Proposition 2.3.5.

Solution. TODO

2.4 Continuous Random Variable

DEFINITION 2.4.1: Continuous random variable

Suppose X is a random variable with c.d.f. F. If F is a continuous function for all $x \in \mathbb{R}$ and F is differentiable except possibly at countably many points, then X is called a **continuous random variable**.

Note that this is not a rigorous definition, but it will be used in this course.

DEFINITION 2.4.2: Probability function, Support set

The **probability function** of a continuous random variable is

$$f(x) = \begin{cases} F'(x) & \text{if } F(x) \text{ is differentiable at } x \\ 0 & \text{otherwise} \end{cases}$$

The set $A = \{x : f(x) > 0\}$ is called the **support set** of X.

Continuous case: $f(x) \neq P(X = x)$

$$P(x < X \leqslant x + \delta) \approx f(x)\delta$$

since

$$\lim_{\delta \to 0} \frac{F(x+\delta) - F(x)}{\delta} = F'(x) = f(x)$$

PROPOSITION 2.4.3: Properties of the Probability Function

(1) $f(x) \geqslant 0$ *for all* $x \in \mathbb{R}$

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = \lim_{x \to \infty} F(x) - \lim_{x \to -\infty} F(x) = 1$

(3) $f(x) = F'(x) = \displaystyle\lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$

(4) $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,dt$ *since* $F(-\infty) = 0$.

(5) $P(a < X \leqslant b) = P(X \leqslant b) - P(X \leqslant a) = F(b) - F(a) = \displaystyle\int_a^b f(x)\,dx$

(6) $P(X = b) = F(b) - \displaystyle\lim_{a \to b^-} F(a) = F(b) - F(b) = 0 \neq f(b)$ *since* F *is continuous.*

EXAMPLE 2.4.4

Suppose the c.d.f. of X is

$$F(x) = \begin{cases} 0 & x \leqslant a \\ \dfrac{x-a}{b-a} & a < x < b \\ 1 & x \geqslant b \end{cases}$$

Find the p.d.f. of X.

Solution.

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

We note that $X \sim \text{Uniform}(a, b)$

EXAMPLE 2.4.5

Let the p.d.f. be defined as follows.

$$f(x) = \begin{cases} \dfrac{\theta}{x^{\theta+1}} & x \geqslant 1 \\ 0 & x < 1 \end{cases}$$

(i) For what values of θ is f a p.d.f.
(ii) Find $F(x)$.
(iii) Find $P(-2 < X < 3)$.

Solution.

(i) Note that $\dfrac{\theta}{x^{\theta+1}} \geqslant 0$ for all $\theta \geqslant 0$.
Case 1: $\theta = 0$. $f(x) \equiv 0$, then f cannot be a pdf since $\int_{-\infty}^{\infty} f(x)\,dx = 0 \neq 1$
Case 2: $\theta > 0$.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx = \int_{1}^{\infty} \frac{\theta}{x^{\theta+1}}\,dx = \left[-x^{-\theta}\right]_1^{\infty} = 1$$

Therefore, f is a p.d.f. when $\theta > 0$.

(ii) $F(x) = P(X \leqslant x)$.
Case 1: $x < 1$.

$$P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt = 0$$

Case 2: $x \geqslant 1$.

$$P(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{1} f(t)\,dt + \int_{1}^{x} f(t)\,dt = \int_{1}^{x} \frac{\theta}{t^{\theta+1}}\,dt = \left[-t^{-\theta}\right]_1^{x} = 1 - x^{-\theta}$$

(iii) $P(-2 < X < 3)$. Either use the c.d.f. we found or the p.d.f.
Using the c.d.f. we have

$$F(3) - F(-2) = (1 - 3^{-\theta}) - 0$$

Using the p.d.f. we have

$$\int_{-2}^{3} f(x)\,dx = \int_{-2}^{1} f(x)\,dx + \int_{1}^{3} f(x)\,dx = \int_{1}^{3} f(x)\,dx = \text{exercise}$$