Introduction to probability - MA2110

Amit Tripathi

Indian Institute of Technology, Hyderabad

July 29, 2019

References

Some references for this course:

1. A First Course in Probability - Sheldon Ross.
2. Introductory Probability and Statistical Applications - Paul Meyer.
3. Elementary Probability Theory - Chung.

Marking etc

- Weekly take home quizzes - $4 \times 5\% = 20\%$.
- One class test - 10%.
- Final exam - 70%.
- Test and Exam will have subjective type questions. Quizzes will be objective type.
- Email: amittr@iith.ac.in

Final exam and Test dates

1. **Final exam is on 31st August (Saturday) from 14:00 - 16:00.**
2. **Test is on 14th August during class time.**

Why is probability theory important?

In mathematics, probability theory is useful to solve (otherwise intractable) problems in

1. Number theory,
2. Combinatorics,
3. Graph Theory and
4. Statistics etc

Outside mathematics, it has applications in

1. Weather predictions.
2. Machine learning.
3. Risk assessment in business.
4. Insurance sector.
5. Stock Market etc

Probability applied to study prime numbers

Question

Give an arithmetic progression of length 3 consisting entirely of prime numbers?

Question

Give an arithmetic progression of length 5 consisting entirely of prime numbers?

What about length 6,7,$\cdots$? Best bound using computers is $k = 26$:

$$43,142,746,595,714,191 + 23,681,770 \cdot 223,092,870 \cdot n$$

For $n = 0, 1, \cdots, 25$.

Probability applied to study prime numbers

Question

Given any k, can we always find an arithmetic progression of length k consisting entirely of prime numbers?

Theorem of Ben Green and Terence Tao - Yes!

Proof uses probability theory in a non-trivial manner.

Experiment

Definition

An experiment or trial is a procedure which can be repeated as often as we like and its set of possible outcomes are known in advance.

1. Outcome of a toss.
2. Runs scored in a given cricket match.
3. Temperature at some place on a given day.
4. Price of a commodity on a given day.

Sample Space

Definition

The set of all possible outcomes of an experiment is called the **sample space** and is denoted usually by S or Ω.

There may be more than one sample space associated with an experiment. Suppose we toss a coin twice:

1. If we are interested in the outcome of two tosses then the sample space is $S = \{HH, TH, HT, TT\}$.
2. If we are interested in the number of heads in two tosses then the sample space is $S = \{0, 1, 2\}$.

Example

Suppose a dice is thrown n times, then one possible sample space is

$$S = \{1, 2, 3, 4, 5, 6\}^n$$

Event

Definition

An event E is any subset of S i.e. $E \subseteq S$.

1. Two events E, F are **disjoint** or **mutually exclusive** if $E \cap F = \emptyset$.
2. A collection of events $\{E_1, \ldots\}$ is **exhaustive** if $\bigcup_i E_i = S$.

Question

What is the difference between an outcome and an event?

Events

Notation:

1. The **complement of** E is written as E^c.
2. Given events E, F, the **union of** E and F is the event $E \cup F$.
3. Given events E, F, the **intersection of** E, F is the event $E \cap F$ (usually written as EF).

Events

Example

1. Suppose we toss a coin thrice. Set of all possible outcomes (sample space) is
$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$
2. Any point in this space is an outcome e.g. HHH is one possible outcome.
3. The event E that *"at least two heads appear"* is $E = \{HHH, HHT, HTH, THH\}$. Similarly the event that *"at least two tails appear"* is $F = \{HTT, THT, TTH, TTT\}$.
4. Clearly E and F are **mutually exclusive** or **disjoint**.
5. Also $\{E, F\}$ are **exhaustive**.

Axioms of Probability

[**Kolmogorov (1933)**] - Let S be the sample space for an experiment. We say that P **describes a probability on** S, if for any event $E \subseteq S$ there exists a number $P(E)$ satisfying following axioms:

1. $P(E) \geq 0$ for all $E \subseteq S$.
2. $P(S) = 1$.
3. Let $E_j \subseteq S, j = 1, \ldots,$ be a disjoint set of events i.e. $E_j \cap E_k = \emptyset$ when $j \neq k$. Then

$$P(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$$

The number $P(E)$ is called the **probability of the event** E.

Axioms of Probability

1. $P(E) \geq 0$ for all $E \subseteq S$.
2. $P(S) = 1$.
3. Let $E_j \subseteq S, j = 1, \ldots,$ be a disjoint set of events i.e. $E_j \cap E_k = \emptyset$ when $j \neq k$. Then

$$P(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$$

Question

What is $P(\emptyset)$? Show that $P(E) \leq 1$ for all events. Show that $P(E^c) = 1 - P(E)$.

Axioms of Probability

1. $P(E) \geq 0$ for all $E \subseteq S$.
2. $P(S) = 1$.
3. Let $E_j \subseteq S, j = 1, \ldots,$ be a disjoint set of events i.e. $E_j \cap E_k = \emptyset$ when $j \neq k$. Then

$$P(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$$

Question

1. If $E \subseteq F$ then show that $P(E) \leq P(F)$.
2. For any two events, show that $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Axioms of Probability

1. $P(E) \geq 0$ for all $E \subseteq S$.
2. $P(S) = 1$.
3. Let $E_j \subseteq S, j = 1, \ldots,$ be a disjoint set of events i.e. $E_j \cap E_k = \emptyset$ when $j \neq k$. Then

$$P(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} P(E_j)$$

Question

Suppose the experiment consists of choosing a real number uniformly from the interval $(0, 1)$. What is the probability that a rational number is picked?

Axioms of probability - example

Example

We toss a **fair** coin twice. Sample space is

$$S = \{HH, HT, TH, TT\}$$

Set consisting of all events is the power set of S and has $2^4 = 16$ elements with $P(S) = 1$ and $P(\emptyset) = 0$. For every other event $0 < P(E) < 1$.
We observe that $S = HH \cup HT \cup TH \cup TT$ and they are mutually exclusive events. Since the coin is fair, all outcomes are equally likely i.e. $P(HH) = P(HT) = P(TH) = P(TT)$. By third axiom

$$P(HH) + P(HT) + P(TH) + P(TT) = P(S) = 1$$

So every outcome has probability $1/4$.

More generally - if the sample space is a finite set and every outcome is equally likely then probability of an event E is given by the classical formula

$$P(E) = \frac{|E|}{|S|}$$

Some inequalities

Lemma

Let $E_1, \cdots, E_n$ be any events. Then

1. **Boole's inequality:** $P(\cup_{r=1}^{n} E_r) \leq \Sigma_{r=1}^{n} P(E_r)$.
2. *As a corollary we deduce -* **Bonferroni's inequality:** $P(\cap_{r=1}^{n} E_r) \geq \Sigma_{r=1}^{n} P(E_r) - (n-1)$.

Exercise

Let $E_1, \cdots, E_n$ be events. Show that

$$
\begin{aligned}
P(\cup_{r=1}^n E_r) =& \Sigma_{r=1}^n P(E_r) - \Sigma_{i<j} P(E_i E_j) + \ldots + \\
& + (-1)^{r+1} \Sigma_{i_1 < \cdots < i_r} P(E_{i_1} \cdots E_{i_r}) \\
& + \ldots + (-1)^{n+1} P(E_1 \cdots E_n).
\end{aligned}
$$

Principle of Inclusion and Exclusion

Example

N persons named x_i $(i = 1, \ldots, N)$ are asked to sit on chairs numbered $i = 1, \ldots, N$. Find the probability that each i, x_i doesn't sits on the chair numbered i. Define events:

$$E_i = \{x_i \text{ sits on the chair } i\}.$$

In terms of E_i, what is the event we are looking at? We need to know the values of $P(E_{i_1} \cdots E_{i_k})$. The event $E_{i_1} \cdots E_{i_k}$ means that we are left with $N - k$ positions free which can be arranged in $(N-k)!$ ways. Thus

$$P(E_{i_1} \cdots E_{i_k}) = \frac{(N-k)!}{N!}$$

Now apply inclusion-exclusion!

Conditional probability

Suppose A and B are two events in a sample space S and we are asked to compute the probability of A given that B has already happens. This is the notion of conditional probability written as $P(A|B)$.

Conditional probability

Example

Suppose our experiment consists of rolling two dice. Then the sample space is $S = \{(i, j)\}$ where $1 \leq i, j \leq 6$.

1. Event A is that the first dice shows 4.
2. Event B is that the sum of two dice is 9.

$P(A) = 6/36 = 1/6$. But if we are given the **additional information that B has already happened**. Then?

Clearly then possibilities for A are given by the event $A \cap B$ and total possibilities are $|B| = 4$. Thus

$$P(A\,|B) = \frac{|A \cap B|}{|B|} = 1/4.$$

Conditional probability

Based on this example we define

Definition

The **conditional probability of event** A **given that** B has already occured is

$$P(A|B) = \frac{P(AB)}{P(B)} \quad \text{where } P(B) \neq 0$$

Conditional probability

Exercise

Conditional probability is a probability.

Independent events

Definition

We say an event E is independent of another event F if knowing that F has happened doesn't affect the probability of E i.e. $P(E|F) = P(E)$. Using the definition of conditional probability this is equivalent to saying

$$P(EF) = P(E)P(F)$$

We take the latter formula as the definition of independence of two events. The former formula $P(E|F) = P(E)$ implicitly assumes that $P(F) = 0$ but latter one is free of this assumption.

Independent events

Example

Consider an arbitrary n digit binary number. Define events

$$H = \{\text{number has both 0 and 1 in its binary representation}\}$$

and

$$A = \{\text{there is at most one 1}\}$$

Then $P(H) = 1 - \dfrac{1}{2^{n-1}}$, and $P(A) = \dfrac{1+n}{2^n}$.
The event AH represent precisely one 1 and rest 0s and thus

$$P(AH) = \frac{n}{2^n}$$

Are A and H independent events?

Independent events

Example

$$P(A)P(H) = \frac{2^{n-1}-1}{2^{n-1}}\frac{n+1}{2^n} >=< \frac{n}{2^n} = P(AH)$$

$$\iff \frac{2^{n-1}-1}{2^{n-1}} >=< \frac{n}{n+1}$$

$$\iff 2^{n-1} >=< n+1$$

Thus we get

$$P(A)P(H) \begin{cases} < P(AH), & n=2 \\ = P(AH), & n=3 \\ > P(AH), & n>3 \end{cases}$$

In particular, when $n=3$ then these two events are independent!

Multiplication Rule

Lemma

Let E_1, E_2 be events then

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

Lemma (Generalized Multiplication Rule)

Let $E_1, \cdots, E_n$ be events then

$$P(E_1E_2\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

Multiplication Rule

Example

A box contains r red balls and w white balls. We draw k balls from the box **without replacement**. Find the probability that all are red balls?

One way is to define event E_i as ith ball is red. Then we are interested in $P(E_1 E_2 \cdots E_k)$. We use **multiplication rule**,

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \cdots P(E_n|E_1 \cdots E_{n-1})$$

Thus we get

$$P(E_1 E_2 \cdots E_k) = \frac{r(r-1)\cdots(r-k+2)(r-k+1)}{(r+w)(r+w-1)\cdots(r+w-k+1)}$$

Law of total probability

Theorem

Let E and F be two events. Then

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$

Corollary

Suppose $S = F_1 \cup F_2 \cup \cdots \cup F_k$ where F_i are mutually disjoint events. Then

$$P(E) = \sum_i P(E|F_i)P(F_i)$$

Law of total probability

Example

Insurance companies classify people in two categories - Health wise high risk and low risk. Probability that a high risk person will have some medical emergency in a year is p and the probability that a low risk person will have some medical emergency in a year is q.

Assume that in a population the probability that a random person is health wise high risk is r.

Question - what is the probability that a new policy holder will have some medical emergency within a year of purchasing a policy?

Law of total probability

Example

High risk will have some medical emergency $= p$, Low risk will have some medical emergency $= q$, a random person is high risk $= r$.

Question - what is the probability that a new policy holder will have some medical emergency within a year of purchasing a policy?

Let E be the event that a new policy holder will have some medical emergency within a year of purchasing a policy. Let A be the event that policy holder is high risk. We don't know $P(E)$ but we know $P(E|A) = p$ and also $P(E|A^c) = q$. Using Law of total probability:

$$P(E) = P(E|A)P(A) + P(E|A^c)(1 - P(A)) = pr + q(1 - r)$$

Bayes' Theorem

Theorem (Bayes)

Suppose E, F be two events. Then

$$P(F|E) = \frac{P(EF)}{P(E)} = \frac{P(E|F)P(F)}{P(E)}$$

Bayes' Theorem

Example

High risk will have some medical emergency $= p$, Low risk will have some medical emergency $= q$, a random person is high risk $= r$.
Question - A policy holder has some medical emergency within one year of purchasing a policy. What is the probability that he/she was high risk? Recall:

$$E = \{\text{a new policy holder will have some medical emergency}\}$$
$$A = \{\text{a new policy holder is high risk}\}$$

We are looking for $P(A|E)$. We already calculated $P(E) = pr + q(1-r)$. Therefore

$$P(A|E) = P(AE)/P(E) = P(E|A)P(A)/P(E) = \frac{pr}{pr + q(1-r)}$$

Bayes' Theorem

Theorem (General version of Bayes' theorem)

Suppose $S = F_1 \cup F_2 \cup \cdots \cup F_k$ *where* F_i *are mutually disjoint events. Then*

$$P(F_i|E) = \frac{P(EF_i)}{P(E)} = \frac{P(EF_i)}{\sum_i P(E|F_i)P(F_i)} = \frac{P(E|F_i)P(F_i)}{\sum_i P(E|F_i)P(F_i)}$$

Random variables

Definition

A random variable $X : S \to \mathbb{R}$ is a real valued function on the sample space.

Example

Suppose we roll a dice twice. Sample space

$$S = \{(i, j) \mid 1 \leq i, j \leq 6\}$$

Let X denote the random variable that is equal to the sum of both dices. Thus, if the outcome is $(1, 2)$ then $X(1, 2) = 3$.

Random variables - More examples

1. Sample space

$$S = \{\text{List of cricket players who played in CWC 2019}\}$$

   Define $X : S \to \mathbb{R}$ where $X(S)$ is the total number of runs scored by the player S.

2. Define $Y : S \to \mathbb{R}$ where $Y(\omega)$ is the total number of wickets taken by the player ω.
3. Define a third random variable $Z : S \to \mathbb{R}$ as

$$Z(\omega) = \lambda X(\omega) + \mu Y(\omega)$$

   as some measure of all-round performance by the player ω where λ and μ are some real numbers (depending upon the model).

Random variables

1. A random variable takes an outcome ω and gives back a real number $X(\omega)$.
2. If X, Y are random variables, then clearly $X \pm Y$ are also random variables with $(X \pm Y)(\omega) = X(\omega) \pm Y(\omega)$.
3. For any $a \in \mathbb{R}$, $\{X = a\}$ denotes the event $\{\omega \,|\, X(\omega) = a\}$.
4. Similarly, $\{X \leq a\}$ will denote the event

$$\{\omega \in S \,|\, X(\omega) \leq a\}$$

Random variables

Quick questions -

1. What is $P(\{X \leq a\}) + P(\{X > a\})$?
2. Toss a fair coin 5 times. Sample space is

$$S = \{HHHHH, HHHHT, HHHTH, HHHTT, \cdots\}$$

For any outcome ω, define $X(\omega)$ to be the number of H (heads) in ω.

    1. What is $P(X = 1)$?
    2. What is $P(X \geq 1)$?
    3. What is $P(1 \leq X \leq 4)$?

Probability Mass Function

Definition

The **probability mass function or PMF** of X is defined as

$$MF_X(a) = P(X = a)$$

Question

If the range of X is countable (say) $\{a_1, \cdots, \}$, then

$$\sum_{i=1} MF_X(a_i) = ?$$

Discrete Random variables

Definition

If the set $X(S) \subset \mathbb{R}$ is countable then X is said to be discrete random variable (DRV).

Example

Number of heads in n tosses of a coin, out come of a throw of dice etc are all DRVs.

Question

If S is countable, does it mean that X is discrete?

Question

If X is discrete, does it mean that S is countable?

Cumulative Distribution Function

Definition

The **cumulative distribution function or CDF** of X is defined as

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

Cumulative Distribution Function

Question

Is CDF always monotonic?

Question

Is CDF left continuous? right continuous?

Question

What is $F_X(\infty)$? What is $F_X(-\infty)$?

Fact - Any non-decreasing, right continuous function, satisfying $F(-\infty) = 0$ and $F(\infty) = 1$, is CDF of some random variable.

Some questions

Question

If the probability mass function (of a DRV) is

$$MF_X(k) = \frac{c\lambda^k}{k!} \quad \text{for} \quad k = 0, 1, 2, \ldots$$

Find c (in terms of λ)?

Question

Suppose CDF of X is F_X. Express $P(a < X \leq b)$ in terms of F_X ?

Discrete Random variables

Definition

If the set $X(S) \subset \mathbb{R}$ is finite or countably infinite then X is said to be discrete random variable (DRV).

Example

Number of heads in n tosses of a coin, out come of a throw of dice etc are all DRVs.

Question

If S is countably infinite, does it mean that X is discrete?

Question

If X is discrete, does it mean that S is finite or countably infinite?

Probability Mass Function

Definition

The **probability mass function or PMF** of X is defined as

$$MF_X(a) = P(X = a)$$

Question

If the range of X is countable (say) $\{a_1, \cdots, \}$, then

$$\sum_{i=1} MF_X(a_i) = ?$$

A question

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2} + \frac{x}{2}, & 0 < x < 1 \\ 1, & 1 \leq x \end{cases}$$

Does F describes cumulative distribution function of some random variable?

Cumulative Distribution Function

Trivial fact - Cumulative distribution function is always non-decreasing, right continuous and satisfies $F(-\infty) = 0$ and $F(\infty) = 1$.

(Not so important but interesting) Fact - Conversely, any non-decreasing, right continuous function, satisfying $F(-\infty) = 0$ and $F(\infty) = 1$, is CDF of some random variable.

Some questions

Question

If the probability mass function (of a DRV) is

$$MF_X(k) = \begin{cases} \dfrac{c\lambda^k}{k!}, & \text{for} \quad k = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

Find c (in terms of λ)?

Some questions

Question

Suppose CDF of X is F_X. Express $P(a < X \leq b)$ in terms of F_X ?

Let X be a random variable. Define a random variable $|X|$ as follows

$$|X| = \begin{cases} X, & X \geq 0 \\ -X, & X < 0 \end{cases}$$

Then $|X|$ is a RV with distribution function given as

$$F_{|X|}(x) = P(|X| \leq x) = \begin{cases} F_X(x) - F_X(-x) + P(X = -x), & x > 0 \\ P(X = 0), & x = 0 \\ 0, & x < 0 \end{cases}$$

Expectation

Definition

Expected value of a discrete RV with probability mass function $MF_X(k)$ then the **expected value** of X denoted by $E[X]$ is defined by

$$E[X] = \sum_{k:MF_X(k)>0} kMF_X(k) = \sum_{k:MF_X(k)>0} k \cdot P(X = k)$$

Think of it as the weighted average of the possible values that X takes (weight assigned according to the probability of X being equal to that value).

An example

Example

A man buys a lottery priced at r with winning prize worth R. Suppose his probability of winning is p. Let X denotes the DRV representing his earnings. Then

$$X \in \{R - r, -r\}$$

$P(X = -r) = 1 - p$ and $P(X = R - r) = p$. Thus

$$E[X] = p(R - r) + (1 - p)(-r) = pR - r$$

Another example

Example

Suppose we flip a coin which has a probability p of coming up heads. We keep flipping it until either a head comes or up to n trials. Let X be the number of times we have to flip it. Then

$$
\begin{aligned}
P(X=1) =& p, P(X=2) = (1-p)p \cdots \\
P(X=k) =& (1-p)^{k-1}p
\end{aligned}
$$

for $k < n$ and $P(X=n) = (1-p)^{n-1}p + (1-p)^n$.
Thus

$$E[X] = \sum_{i=1}^{n} i(1-p)^{i-1}p + n(1-p)^n = \frac{1-(1-p)^n}{p}$$

Can the expectation of a finite valued random variable be infinite?

Example

Let probability mass function of a random variable be given as

$$P(X = m) = \begin{cases} \frac{\lambda}{m^2}, & m \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

By a Theorem of Euler, $\lambda = \frac{6}{\pi^2}$.

$$E[X] = \sum_{m>0} mP(X = m) = \sum_{m>0} \frac{\lambda}{m} = \infty$$

Expectation - alternate viewpoint

$$E[X] = \sum_{\omega \in S} X(\omega)p(\omega)$$

Thus X is the **weighted average** as X ranges over outcomes in S.

Expectation - alternate viewpoint

Proof.

Suppose $X(S) = \{x_1, \cdots\}$. Let $E_i = \{\omega \mid X(\omega) = x_i\}$. In particular

$$S = \sqcup_i E_i$$

By definition

$$\begin{aligned} E[X] &= \sum_i x_i P(X = x_i) \\ &= \sum_i x_i \sum_{\omega \in E_i} p(\omega) \\ &= \sum_i \sum_{\omega \in E_i} X(\omega) p(\omega) \\ &= \sum_{\omega \in S} X(\omega) p(\omega) \end{aligned}$$

□

Linearity of expectation

Lemma

$E[X+Y] = E[X] + E[Y]$

We use weighted average interpretation of expectation - Suppose $Z = X + Y$. Then

$$\begin{aligned} E[Z] &= \sum_{\omega \in S} Z(\omega) p(\omega) \\ &= \sum_{\omega \in S} (X(\omega) + Y(\omega)) p(\omega) \\ &= \sum_{\omega \in S} X(\omega) p(\omega) + \sum_{\omega \in S} Y(\omega) p(\omega) \\ &= E[X] + E[Y] \end{aligned}$$

Indicator random variable

Definition

Let A be any event. We define the **indicator random variable** I_A of A as

$$I_A = \begin{cases} 1, & A \text{ happens} \\ 0, & A \text{ doesn't happens} \end{cases}$$

What is expected value of I_A ?

Then $E[I_A] = 0 \cdot P(A \text{ doesn't happens}) + 1 \cdot P(A \text{ happens}) = P(A)$.

Variance

Definition

Let X be a RV with mean μ. Then the **variance** of X, denoted $Var(X)$ is defined by

$$Var(X) = E[(X - \mu)^2]$$

Usually denoted as σ_X^2.

Question

Suppose $P(A) = p$ and I_A is indicator random variable. What is $Var(I_A)$?