INTRODUCTORY STATISTICS

STATS101
SINGAPORE MANAGEMENT UNIVERSITY
Tutor: Ravins Cheung (A+)

Module 2: Discrete Distribution Problems

1. Probability (Combinations and types of events)
2. Probability (Classical Probability model)
3. Probability (Rules of probability)
4. Probability (Conditional probability)
5. Discrete random variables
6. Discrete probability distributions
7. Binomial distributions
8. Poisson distribution
9. Recap

ZOOKAL

Probability

Probability of an Event

- The probability of event E is a numerical value which measures the likelihood of occurrence of E.

Note

- $0 \leq P(E) \leq 1$ where $P(E) =$ probability of event E.
- $P(E) = 0$ implies that it is impossible for E to occur.
- $P(E) = 1$ implies that E is certain to occur.
- Likelihood of E occurring is greater the closer $P(E)$ is to 1.
- Likelihood of E occurring may also be assessed by

$$\text{odds of event } E = \frac{P(E)}{1 - P(E)}, \text{ assuming } P(E) \neq 1.$$

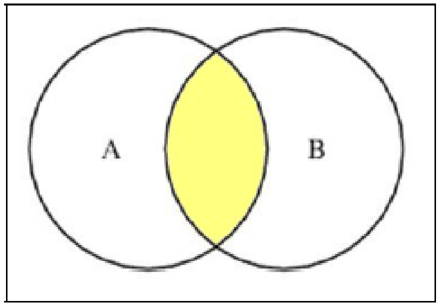
Combinations and types of events

Combinations of Events

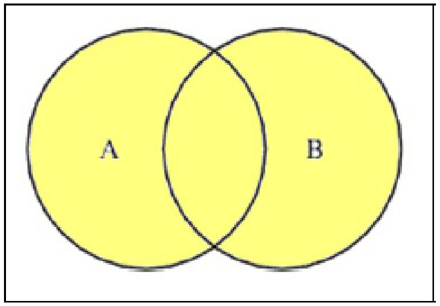
Let A and B denote any two events associated with a random experiment.

- The **intersection of events** A and B, denoted $A \cap B$, is the event that both A and B occurs.
- The **union of events** A and B, denoted $A \cup B$, is the event that at least one of the two events A or B occurs.

$A \cap B$



$A \cup B$



Combinations and types of events

Types of Events

- A and B are **disjoint events** if

$$P(A \cap B) = 0.$$

- A and B are **complementary events** if

$$P(A \cap B) = 0$$

  and

$$P(A \cup B) = 1.$$

- A and B are **independent events** if[1]

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

Classical Probability Model

The CPM Model

- Under CPM, the probability of event E is given by

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

where

$$\begin{aligned} n(E) &= \text{number of outcomes in event } E, \\ n(S) &= \text{number of outcomes in sample space } S. \end{aligned}$$

Note

- The **sample space** S is the set of all possible outcomes.
- An event E is a subset of S.
- CPM assumes S is finite and outcomes in S are equally likely.

Classical Probability Model

A dice is rolled once, what is the chance of getting less than 3?

Set of possible outcomes is,
S = {1,2,3,4,5,6}
*S is a finite set of equally likely outcomes.

The event of interest is when dice is equal to 1 or 2
E = {1,2}

Since n(E) = 2, by CPM

$$P(E) = \frac{2}{6} = .3333$$

Rules of Probability

Let A and B denote two events associated with a particular random experiment.

Law of Complementation (LC)

$$P(\overline{A}) = 1 - P(A)$$

General Addition Rule (GAR)

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

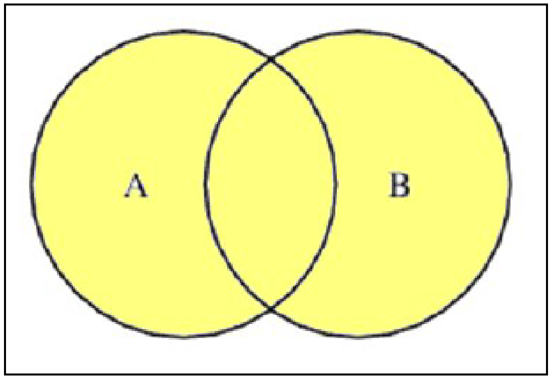
Theorem of Total Probability (TTP)

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

Rules of Probability

Intuition Underlying GAR

From the figure below, we note that $n(A) + n(B)$ overestimates $n(A \cup B)$ by $n(A \cap B)$.



Hence,

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

and this yields GAR on dividing by $n(S)$.

Conditional Probability

- **Conditional Probability**

  The conditional probability of event A given occurrence of event B is defined as[3]

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

  provided $P(B) \neq 0$.

- **Multiplication Rule (MR)**

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

Conditional Probability

Note

- LC, GAR & TTP also applies to conditional probabilities, e.g., the conditional version of LC is: $P(\overline{A}|B) = 1 - P(A|B)$.

  Q: What are the conditional versions of GAR and TTP?
- The following is also valid decomposition

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

- By repeated application of MR, we can also decompose the probability of joint occurrence of more than two events as a product of probabilities, e.g.,

$$\begin{aligned} P(A \cap B \cap C) &= P(A|B \cap C)P(B \cap C) \\ &= P(A|B \cap C)P(B|C)P(C). \end{aligned}$$

  Other decompositions of $P(A \cap B \cap C)$ are also possible.

Conditional Probability

The table reflects the relative frequency of distribution of marital status of adults in a town.

Marital Status	**Male**	**Female**	**Total**
Single	0.18	0.15	**0.33**
Married	0.33	0.13	**0.45**
Others	0.13	0.10	**0.23**
Total	**0.63**	**0.38**	**1.00**

Suppose that an adult selected is at random.

1. What is the probability that the adult selected is single given that the selected person is male?
2. What is the probability that the adult selected is male given that the selected person is single?

ZOOKAL

Conditional Probability

To obtain P(A|B) and P(B|A) where

A = selected adult is single
B = selected adult is male

By random selection, the given table states that,

$$P(A) = 0.3,\ P(B) = 0.63,\ \text{and}\ P(A \cap B) = 0.18$$

Therefore,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.18}{0.63} = 0.2857$$

and

$$P(B|A) = \frac{P(A \cap B)}{P(B)} = \frac{0.18}{0.3} = 0.6$$

ZOOKAL

Conditional Probability

Independent Events

- A and B are (statistically) **independent events** if and only if

$$P(A|B) = P(A),$$

or, equivalently, if and only if

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

Note

- If A and B are independent events, can we conclude the same for the following: (i) $\overline{A}$ and B, (ii) A and $\overline{B}$, (iii) $\overline{A}$ and $\overline{B}$?
- Events $E_1, E_2, \ldots, E_n$ are independent if and only if

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = P(E_{i_1})P(E_{i_2}) \cdots P(E_{i_k}), \quad k = 2, \ldots, n.$$

Conditional Probability

Consider the situation whereby a box contains 3 red balls and 7 blue balls.

A = second item selected is blue
B = first item selected is blue

When sampling without replacement,
$$P(A) = 0.7, P(A|B) = 0.67$$

When sampling with replacement,
$$P(A) = 0.7, P(A|B) = 0.7$$

Thus, A and B are independent events in the second case but not the first.

How would you

Conditional Probability

Bayes Theorem

- If $A_1, \ldots, A_n$ are exhaustive (and disjoint) events, then

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)},$$

for $i = 1, \ldots, n$.

Note

- The $P(A_i)$'s are **prior probabilities**.
- The $P(A_i \mid B)$'s are **posterior probabilities**.
- Bayes theorem allows one to revise prior probabilities of $A_1, \ldots, A_n$ once event B is observed to obtain the corresponding posterior probabilities.

Bayes Theorem

Region	Percentage of Population	Percentage of Children below age 7
North	15	5
South	20	6
East	30	7
West	35	5

Supposed a resident is selected at random. What is the probability that the resident is from the North region, given that he/she is a child.

Consider the following events,

S = selected resident is a Child

R1 = Selected resident is from the North

Priori, where P(R1) = 0.15

Bayes Theorem

Region	Percentage of Population	Percentage of Children below age 7
North	15	5
South	20	6
East	30	7
West	35	5

By Bayes Theorem,

P(R1|S) =

Discrete Random Variables

- Loosely speaking, a **random variable** is one that can take different values in its range with different probabilities.
- X is a **discrete random variable** if its range R_X is finite or countably infinite.

For example, number of red balls in a box of balls, number of times required to roll a dice to get the first “1”.

Events in Terms of Random Variables

An event in terms of random variable X is a statement that X is in a certain subset of its range.

Consider random variable X which counts the number of “1”s in two rolls of a fair dice.

Discrete Probability Distributions

Probability Mass Function (PMF)

- The **probability mass function** of a discrete random variable X is

$$p_X(u) = P(X = u), \quad u \in R_X.$$

The PMF of X, the number of “1”s in two rolls of a fair dice, is

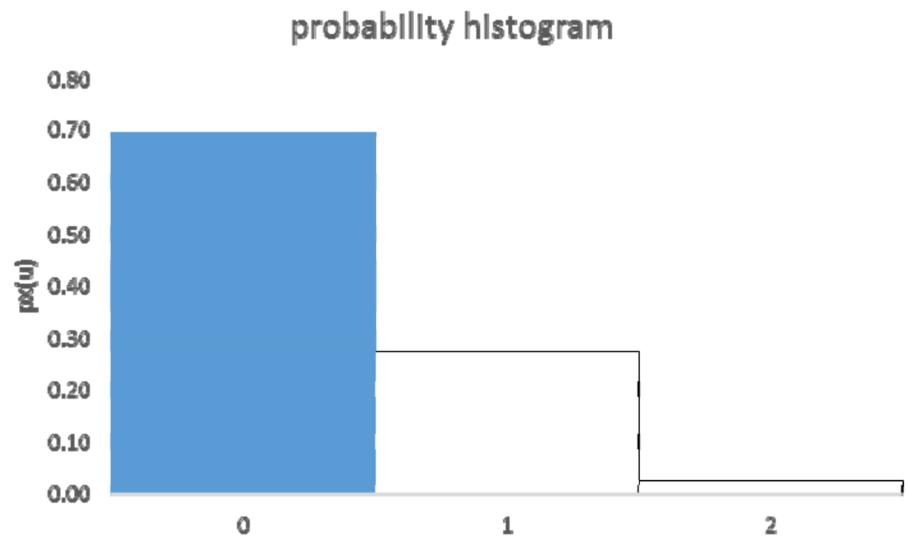
u	0	1	2
px(u)	0.69	0.28	0.03

This follows from (note that from (LC) px(2)=1-px(1)-px(0))
X = 0 ⇔ {(=/=6,=/=6)} occurs,
X = 1 ⇔ {(=6,=/=6), (=/=6,=6)} occurs,

Discrete Probability Distributions

u	0	1	2
px(u)	0.69	0.28	0.03

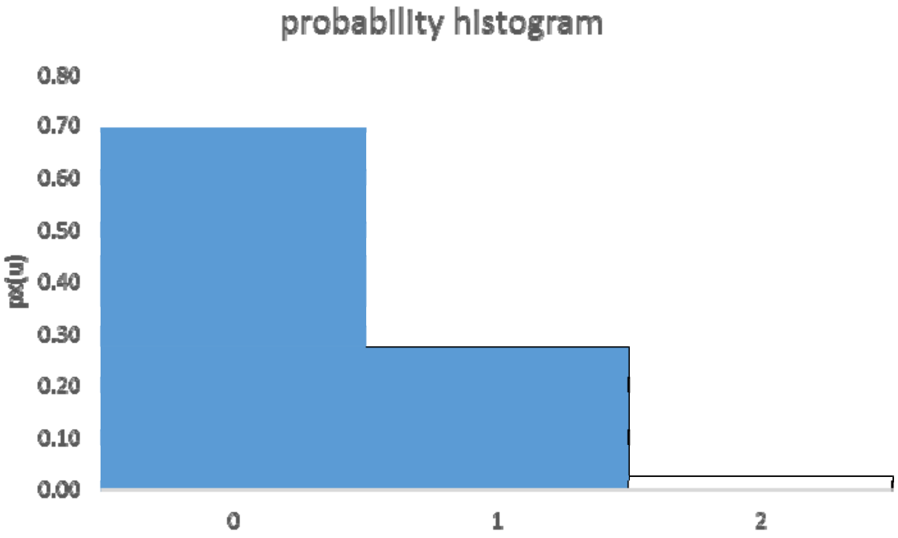
Graphing the probability histogram of the above table, the shaded region corresponds to the respective probability, P(X=0)



Discrete Probability Distributions

u	0	1	2
px(u)	0.69	0.28	0.03

Graphing the probability histogram of the above table, the shaded region corresponds to the cumulative probability, $P(X \leq 1)$



Discrete Probability Distributions

- If a and b (with $a < b$) are in the range of X, then

$$P(a \leq X \leq b) = \sum_{u=a}^{b} p_X(u),$$

assuming X is an *integer-valued* random variable with $p_X(\cdot)$ as the PMF.

Note that for discrete random variables, it matters whether the inequality involved is '$<$' or '$\leq$' as can be seen from:

- $P(a < X < b) = P(a + 1 \leq X \leq b - 1)$,
- $P(a \leq X < b) = P(a \leq X \leq b - 1)$,
- $P(a < X \leq b) = P(a + 1 \leq X \leq b)$.

Discrete Probability Distributions

u	0	1	2
px(u)	0.69	0.28	0.03

$P(0<X<2) =$

$P(0 \leq X<2) =$

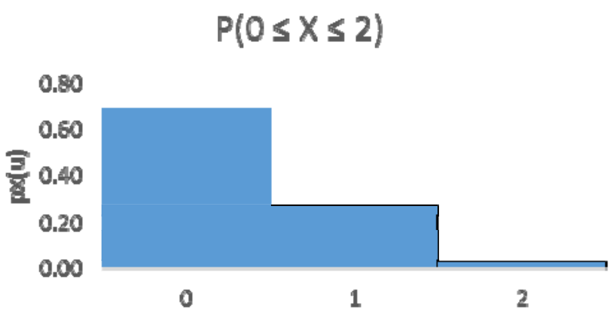
$P(0<X \leq 2) =$

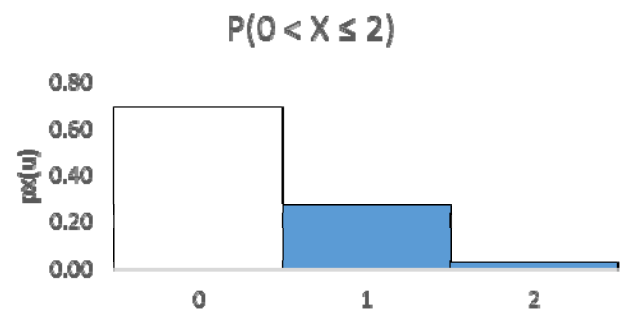
$P(0 \leq X \leq 2) =$

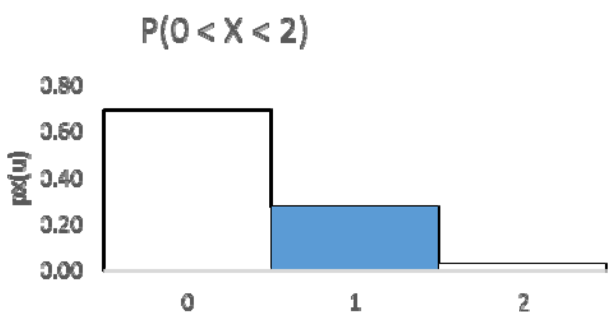
Discrete Probability Distributions

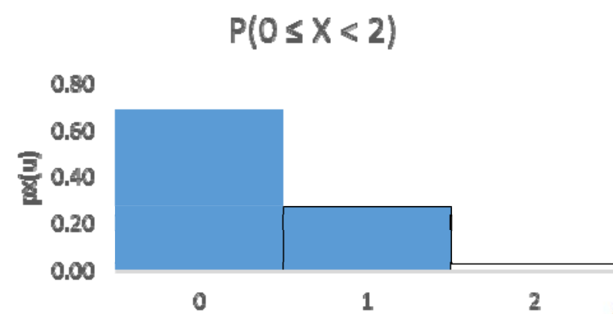
u	0	1	2
px(u)	0.69	0.28	0.03

Probability histograms









ZOOKAL

Discrete Random Variables

Mean & Variance for a Discrete Random Variable

Let $R_X = \{x_1, x_2, \ldots, x_N\}$ denote the range of discrete random variable X (note that, here, R_X is assumed to be finite).

- The mean of X is

$$E(X) = \sum_{i=1}^{N} x_i P(X = x_i) = \sum_{i=1}^{N} x_i p_X(x_i) = \mu, \text{ say.}$$

- The variance of X is

$$\begin{aligned} var(X) = E\left((X-\mu)^2\right) &= \sum_{i=1}^{N} (x_i - \mu)^2 P(X = x_i) \\ &= \sum_{i=1}^{N} x_i^2 p_X(x_i) - \mu^2 = \sigma^2, \text{ say.} \end{aligned}$$

Discrete Random Variables

u	px(u)	u*px(u)	u^2 *px(u)
0	0.69	0	0
1	0.28	0.278	0.278
2	0.03	0.056	0.111
		0.334	0.389

$E(X) = 0.334$

$Var(X) = 0.389 - 0.334^2 = 0.278$

Binomial Distribution

Bernoulli Trial

- A **Bernoulli trial** is a random experiment with two outcomes which we refer to (generically) as 'Success' and 'Failure'.

Binomial Distribution

- A random variable defined by

  $X =$ number of 'Successes' in n independent Bernoulli trials

  has a **binomial distribution** with PMF

  $$p_X(u) = \binom{n}{u} p^u (1-p)^{n-u}, \qquad u = 0, 1, \ldots, n$$

  if

  $p =$ probability of a 'Success' in a single trial

  is the same for each trial.

Note

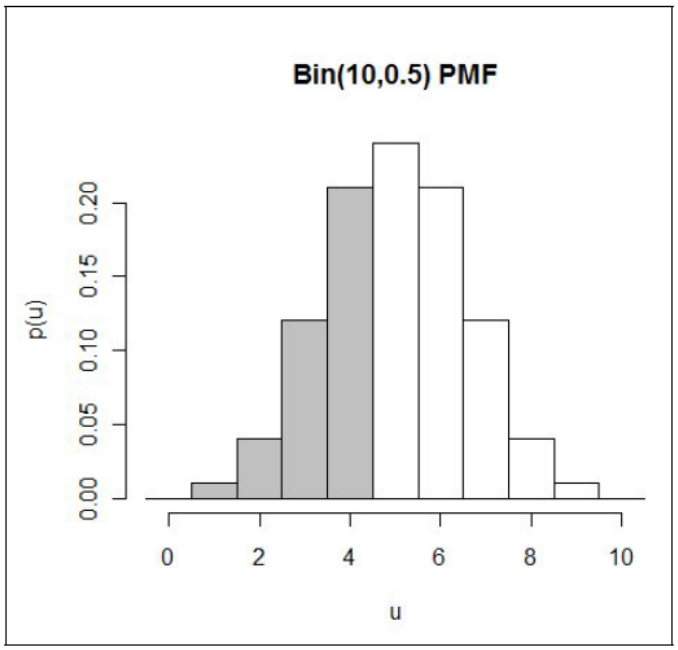
- When X has this distribution, we express this by writing

$$X \sim Bin(n, p).$$

Binomial Distribution

Cumulative Binomial Probability

When $X \sim Bin(10, 0.5)$, the probability $P(X \leq 4)$ is represented by the shaded portion of the corresponding probability histogram.



Binomial Distribution

Consider the following problem where,

X = number of children out of 3 above age 3

In a random sample of three children below age 7,
Assuming $X \sim Bin(3, 0.45)$.

$$P(X \leq 1) = P(\text{At most 1 above age 3})$$
$$= P(X=0) + P(X=1)$$
$$= 0.104$$

And

$$P(X \geq 1) = P(\text{At least one above age 3})$$
$$= 1 - P(X=0) \text{ , by LC}$$
$$0.992$$

ZOOKAL

Binomial Distribution

Mean & Variance of $Bin(n, p)$

- When $X \sim Bin(n, p)$, we have

$$E(X) = np \text{ and } var(X) = np(1 - p).$$

Using the previous example,

E(X) = 3 x 0.8 = 2.4, var(X) = 3 x 0.8 – (1-0.8) = 0.48

Poisson Distribution

PMF, Mean & Variance of a Poisson Distribution

- The PMF of a Poisson distribution is

$$p_X(u) = \frac{e^{-\lambda}\lambda^u}{u!}, \quad u = 0, 1, 2, 3, \ldots$$

- The mean and variance of a Poisson distribution is

$$E(X) = \lambda \text{ and } var(X) = \lambda.$$

Note

- When X has this distribution, we express this by writing

$$X \sim Pois(\lambda).$$

Poisson Distribution

Let X denote the number of customer arrivals at a restaurant between 3:00pm and 4:00pm.

Past records suggests that X ~ Pois(6.9), whereby PMF of X is given by,

$$\text{Px(u)} = \frac{e^{-6.9}6.9^u}{u!}, \text{ u} = 0,1,2,3,\ldots$$

The probability of 4 arrivals between 6:00pm and 7:00pm is

$$\text{P(X=4)} = \text{Px(4)} = \frac{e^{-6.9}6.9^4}{4!} = 0.095$$

And the probability of at most 2 arrivals is

$$\text{P}(X \leq 2) = \text{Px(0)} + \text{Px(1)} + \text{Px(2)} = 0.032$$

Poisson Approximation to Binomial

Poisson Approximation to Binomial

- Suppose $X \sim Bin(n, p)$ with $n \geq 100$ and $np \leq 10$. Then

$$P(X = x) \approx P(Y = x), \quad Y \sim Pois(np).$$

Suppose X~Bin(500,0.0035) where X is the number of rotten apples out of 500 randomly selected picked apples.

Note that, n=500 and np=1.75,

$$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$$
$$\approx P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3), Y \sim Pois(1.75)$$
$$= 0.899$$

Recap

Module 2: Discrete Distribution Problems

1. Probability (Combinations and types of events)
2. Probability (Classical Probability model)
3. Probability (Rules of probability)
4. Probability (Conditional probability)
5. Discrete random variables
6. Discrete probability distributions
7. Binomial distributions
8. Poisson distribution
9. Recap

Next Module:
Module 3: Joint Distribution Problems