Lecture 4

Random Variables
Probability Distributions

Index

- Random Variable
- Bernoulli Random Variable
- Probability Distribution
- Parameter
- Cumulative Distribution
- Expectation

Random Variable

Random Variable is just a variable/rule which is going to assign numerical values to each outcome of sample space.
Random variables are denoted by uppercase letters, such as X and Y.
The notation $X(s) = x$ means that x is the value associated with the outcome s by the rv X.

Example

When a student calls a university help desk for technical support, he/she will either immediately be able to speak to someone (S, for success) or will be placed on hold (F, for failure).

With Sample Space = {S,F}, define an rv X by

X(S) = 1 and X(F) =0

The rv X indicates whether (1) or not (0) the student can immediately speak to someone.

Bernoulli Random Variable

Any random variable whose only possible values are 0 and 1 is called a Bernoulli random variable.

Types of Random Variables

Discrete: A discrete variable is a variable whose value is obtained by counting.

Continuous: A continuous variable is a variable whose value is obtained by measuring.

Probability Distribution

The probability distribution or probability mass function (pmf) of a discrete rv is defined for every number x by

$$p(x) = P(X=x)$$

For Example $p(0) = P(X=0)$

The values of X along with their probabilities collectively specify the pmf.

Example

Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

Let X be the number of defectives in the selected lot. The three possible X values are 0, 1, and 2.

$$p(0) = P(X = 0) = P(\text{lot 1 or 3 or 6 is sent}) = \frac{3}{6} = .500$$

$$p(1) = P(X = 1) = P(\text{lot 4 is sent}) = \frac{1}{6} = .167$$

$$p(2) = P(X = 2) = P(\text{lot 2 or 5 is sent}) = \frac{2}{6} = .333$$

Example

Consider whether the next person buying a computer at a certain electronics store buys a laptop or a desktop model. Let

$$X = \begin{cases} 1 & \text{if the customer purchases a desktop computer} \\ 0 & \text{if the customer purchases a laptop computer} \end{cases}$$

If 20% of all purchasers during that week select a desktop, the pmf for X is

$$p(0) = P(X = 0) = P(\text{next customer purchases a laptop model}) = .8$$
$$p(1) = P(X = 1) = P(\text{next customer purchases a desktop model}) = .2$$
$$p(x) = P(X = x) = 0 \text{ for } x \neq 0 \text{ or } 1$$

An equivalent description is

$$p(x) = \begin{cases} .8 & \text{if } x = 0 \\ .2 & \text{if } x = 1 \\ 0 & \text{if } x \neq 0 \text{ or } 1 \end{cases}$$

Question

Consider a group of five potential blood donors—a, b, c, d, and e—of whom only a and b have type O+ blood. Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified. Let the rv Y = number of typings necessary to identify an individual with O+ blood.

Note: Once a donor is selected he cannot be selected again.

Find pmf of Y

Question

$$p(1) = P(Y = 1) = P(a \text{ or } b \text{ typed first}) = \frac{2}{5} = .4$$

$$p(2) = P(Y = 2) = P(c, d, \text{ or } e \text{ first, and then } a \text{ or } b)$$

$$= P(c, d, \text{ or } e \text{ first}) \cdot P(a \text{ or } b \text{ next} \mid c, d, \text{ or } e \text{ first}) = \frac{3}{5} \cdot \frac{2}{4} = .3$$

$$p(3) = P(Y = 3) = P(c, d, \text{ or } e \text{ first and second, and then } a \text{ or } b)$$

$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = .2$$

$$p(4) = P(Y = 4) = P(c, d, \text{ and } e \text{ all done first}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = .1$$

$$p(y) = 0 \quad \text{if } y \neq 1, 2, 3, 4$$

Parameter of Probability Distribution

Suppose p(x) depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a parameter of the distribution. The collection of all probability distributions for different values of the parameter is called a family of probability distributions.

Bernoulli distribution (Each different number α between 0 and 1 determines a different member of the Bernoulli family of distributions.)

$$p(x;\alpha) = \begin{cases} 1-\alpha & \text{if } x = 0 \\ \alpha & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Question

Starting at a fixed time, we observe the gender of each student coming inside the class until a boy (B) comes. Let p = P(B), assume that successive coming of students inside the class are independent, and define the rv X by x = number of students observed. Find out the pmf.

$$p(1) = P(X = 1) = P(B) = p$$
$$p(2) = P(X = 2) = P(GB) = P(G) \cdot P(B) = (1 - p)p$$

and

$$p(3) = P(X = 3) = P(GGB) = P(G) \cdot P(G) \cdot P(B) = (1 - p)^2 p$$

Continuing in this way, a general formula emerges:

$$p(x) = \begin{cases} (1 - p)^{x-1} p & x = 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases}$$

Cumulative Distribution Function

The **cumulative distribution function** (cdf) $F(x)$ of a discrete rv variable X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y)$$

For any number x, $F(x)$ is the probability that the observed value of X will be at most x.

Question

A store carries flash drives with either 1 GB, 2 GB, 4 GB, 8 GB, or 16 GB of memory. The accompanying table gives the distribution Y = the amount of memory in a purchased drive:

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

, F(7.999)

Question

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

$$F(1) = P(Y \le 1) = P(Y = 1) = p(1) = .05$$
$$F(2) = P(Y \le 2) = P(Y = 1 \text{ or } 2) = p(1) + p(2) = .15$$
$$F(4) = P(Y \le 4) = P(Y = 1 \text{ or } 2 \text{ or } 4) = p(1) + p(2) + p(4) = .50$$
$$F(8) = P(Y \le 8) = p(1) + p(2) + p(4) + p(8) = .90$$
$$F(16) = P(Y \le 16) = 1$$
$$F(2.7) = P(Y \le 2.7) = P(Y \le 2) = F(2) = .15$$
$$F(7.999) = P(Y \le 7.999) = P(Y \le 4) = F(4) = .50$$

$$F(y) = \begin{cases} 0 & y < 1 \\ .05 & 1 \le y < 2 \\ .15 & 2 \le y < 4 \\ .50 & 4 \le y < 8 \\ .90 & 8 \le y < 16 \\ 1 & 16 \le y \end{cases}$$

Cumulative Distribution Function

For any two numbers a and b with $a \leq b$,

$$P(a \leq X \leq b) = F(b) - F(a-)$$

where "$a-$" represents the largest possible X value that is strictly less than a. In particular, if the only possible values are integers and if a and b are integers, then

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a \text{ or } a + 1 \text{ or} \ldots \text{ or } b) \\ &= F(b) - F(a - 1) \end{aligned}$$

Taking $a = b$ yields $P(X = a) = F(a) - F(a - 1)$ in this case.

X can take values 0, 1, 2, 3, 4, 5

We need to find $P(2 \le X \le 4)$

$= P(X=2) + P(X=3) + P(X=4) = p(2)+p(3)+p(4)$

Now $F(4) = p(4) + p(3) + p(2) + p(1) + p(0)$

From $F(4)$ we need to subtract

$p(1)$ and $p(0)$ to get $p(2)+p(3)+p(4)$

so $F(4) - p(0) - p(1)$

$= F(4) - F(1)$

Now to find $P(X=4)$

$$
\begin{aligned}
& F(4) = p(4) + p(3) + p(2) + p(1) + p(0) \\
- \; & F(3) = p(3) + p(2) + p(1) + p(0) \\
\hline
& F(4) - F(3) = p(4) = P(X=4)
\end{aligned}
$$

so $P(X=4) = F(4) - F(3)$

Question

A consumer organization that evaluates new automobiles reports the number of major defects in each car examined. Let X denote the number of major defects in a randomly selected car of a certain type. The cdf of X is as follows:

$$F(x) = \begin{cases} 0 & x < 0 \\ .06 & 0 \le x < 1 \\ .19 & 1 \le x < 2 \\ .39 & 2 \le x < 3 \\ .67 & 3 \le x < 4 \\ .92 & 4 \le x < 5 \\ .97 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$$

Calculate the following probabilities directly from the cdf:

a. $p(2)$, that is, $P(X = 2)$ **b.** $P(X > 3)$

c. $P(2 \le X \le 5)$ **d.** $P(2 < X < 5)$

Question

1. $P(X=2) = F(X=2) - F(X=1) = 0.39 - 0.19 = 0.20$
2. $P(X>3) = 1 - P(X \leq 3) = 1 - 0.67 = 0.33$
3. $P(2 \leq X \leq 5) = F(5) - F(1) = 0.97 - 0.19 = 0.78$
4. $P(2 < X < 5) = F(4) - F(2) = 0.92 - 0.39 = 0.53$

Expected Values

Let X be a discrete rv with set of possible values D and pmf $p(x)$. The **expected value** or **mean value** of X, denoted by $E(X)$ or μ_X or just μ, is

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Question

Just after birth, each newborn child is rated on a scale called the Apgar scale. The possible ratings are 0, 1, . . . , 10, with the child's rating determined by color, muscle tone, respiratory effort, heartbeat, and reflex irritability (the best possible score is 10). Let X be the Apgar score of a randomly selected child born at a certain hospital during the next year, and suppose that the pmf of X is

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	.002	.001	.002	.005	.02	.04	.18	.37	.25	.12	.01

Question

Just after birth, each newborn child is rated on a scale called the Apgar scale. The possible ratings are 0, 1, . . . , 10, with the child's rating determined by color, muscle tone, respiratory effort, heartbeat, and reflex irritability (the best possible score is 10). Let X be the Apgar score of a randomly selected child born at a certain hospital during the next year, and suppose that the pmf of X is

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	.002	.001	.002	.005	.02	.04	.18	.37	.25	.12	.01

$$
\begin{aligned}
E(X) = \mu &= 0(.002) + 1(.001) + 2(.002) \\
&\quad + \cdots + 8(.25) + 9(.12) + 10(.01) \\
&= 7.15
\end{aligned}
$$

Question

Let X, the number of interviews a student has prior to getting a job, have pmf

$$p(x) = \begin{cases} k/x^2 & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Hint: Use summation Σ

Question

Let X, the number of interviews a student has prior to getting a job, have pmf

$$p(x) = \begin{cases} k/x^2 & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = E(X) = \sum_{x=1}^{\infty} x \cdot \frac{k}{x^2} = k \sum_{x=1}^{\infty} \frac{1}{x}$$

The Expected Value of a Function

If the rv X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_D h(x) \cdot p(x)$$

A computer store has purchased three computers of a certain type at $500 apiece. It will sell them for $1000 apiece. The manufacturer has agreed to repurchase any computers still unsold after a specified period at $200 apiece. Let X denote the number of computers sold, and suppose that p(0) = 0.1, p(1) = 0.2, p(2) = 0.3 and p(3) = 0.4. h(X) denote the profit associated. h(X) = 800X-900. Calculate the expected profit.

Question

A computer store has purchased three computers of a certain type at \$500 apiece. It will sell them for \$1000 apiece. The manufacturer has agreed to repurchase any computers still unsold after a specified period at \$200 apiece. Let X denote the number of computers sold, and suppose that p(0) = 0.1, p(1) = 0.2, p(2) = 0.3 and p(3) = 0.4. h(X) denote the profit associated. h(X) = 800X-900. Calculate the expected profit.

$$\begin{aligned} E[h(X)] &= h(0) \cdot p(0) + h(1) \cdot p(1) + h(2) \cdot p(2) + h(3) \cdot p(3) \\ &= (-900)(.1) + (-100)(.2) + (700)(.3) + (1500)(.4) \\ &= \$700 \end{aligned}$$

References

Probability and statistics for engineers RA Johnson, I Miller, JE Freund - 2000 - 117.239.47.98

Statistics for business & economics DR Anderson, DJ Sweeney, TA Williams, JD Camm

Probability and statistics for engineering and science J Deovre