of X to be the value of x_n such that $F_X(x_n) = 0.01n$ or $n/_{100}$, that is, the probability that X is smaller than x_n is n%.

Example A random variable X has the pdf given by $f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & elsewhere \end{cases}$ Find the lower, middle and upper quartiles.

Solution

On the interval $0 \le x \le 1$, the cdf of X is given by $F(x) = x^2$ thus

a) At lower quartile Q_1, $F(Q_1) = Q_1^{\ 2} = 0.25 \Rightarrow Q_1 = \sqrt{0.25} = 0.5$

b) At median m, $F(m) = m^2 = 0.5 \Rightarrow m = \sqrt{0.5} = \frac{1}{\sqrt{2}}$

c) At upper quartile u $F(Q_3) = Q_3^{\ 2} = 0.75 \Rightarrow Q_3 = \sqrt{0.75} = \frac{\sqrt{3}}{2}$

Qn Find the 64th percentile of the pdf in the above example.

A continuous r.v X has the pdf given by $f(x) = \begin{cases} k(1-x) & \text{for } 0 \le x \le 1 \\ 0, & elsewhere \end{cases}$, find the of mode.

2. PROBABILITY DISTRIBUTION

2.1 Discrete Distribution

Among the discrete distributions that we will look at includes the Bernoulli, binomial, Poisson, geometric and hyper-geometric

2.1.1 Bernoulli distribution

Definition: A Bernoulli trial is a random experiment in which there are only two possible outcomes - success and failure. Eg

- Tossing a coin and considering heads as success and tails as failure.
- Checking items from a production line: success = not defective, failure = defective.
- Phoning a call centre: success = operator free; failure = no operator free.

A Bernoulli random variable X takes the values 0 and 1 and $P(X=1) = p$ and $P(X=0) = 1-p$

Definition: A r.v X is said to be a real Bernoulli distribution if it's pmf is given by;

$$P(X=x) = \begin{cases} p^x(1-p)^{1-x} & for \;\; x = 0,1 \\ 0 & otherwise \end{cases}$$

We abbreviate this as $X \sim B(p)$ ie p is the only parameter here. It can be easily checked that the mean and variance of a Bernoulli random variable are $\mu = p$ and $\sigma^2 = p(1-p1)$

2.1.2 Binomial Distribution

Consider a sequence of *n* independent, Bernoulli trials, with each trial having two possible outcomes, *success* or *failure*. Let *p* be the probability of a success for any single trial. Let*X* denote the number of successes on *n* trials. The random variable *X* is said to have a **binomial distribution** and has probability mass function

$$P(X=x) = {}_nC_r \times p^x(1-p)^{n-x} \;\; for \;\; x = 0,1,2.....n$$

We abbreviate this as $X \sim \text{Bin}\,(n, \text{p})$ read as “X follows a binomial distribution with parameters n and p”. ${}_nC_r$ Counts the number of outcomes that include exactly x successes and $n-x$ failures.
The mean and variance of a Binomial random variable are respectively given by;

$$\mu = np \quad \text{and} \quad \sigma^2 = np(1-p1)$$

Let’s check to make sure that if X has a binomial distribution, then $\sum_{x=0}^{n} P(X = x) = 1$. We will need the binomial expansion for any polynomial:

$$(p+q)^n = \sum_{x=0}^{n} {}_nC_r \times p^x q^{n-x} \text{ therefore } \sum_{x=0}^{n} {}_nC_x \times p^x (1-p)^{n-x} = [p + (1-p)]^n = 1^n = 1 \text{ so}$$

Example 1
A biased coin is tossed 6 times. The probability of heads on any toss is 0:3. Let X denote the number of heads that come up. Calculate: (i) $P(X = 2)$ (ii) $P(X = 3)$ (iii) $P(1 < X < 5)$
Solution
If we call heads a success then X has a binomial distribution with parameters n=6 and p=0:3.

(i) $P(X = 2) = {}_6C_2 \times (0.3)^2 (0.7)^4 = 0.324135$

(ii) $P(X = 3) = {}_6C_3 \times (0.3)^3 (0.7)^3 = 0.18522$

(iii) $$\begin{aligned} P(1 < X \le 5) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ &= 0.324 + 0.185 + 0.059 + 0.01 = 0.578 \end{aligned}$$

Example 2
A quality control engineer is in charge of testing whether or not 90% of the DVD players produced by his company conform to specifications. To do this, the engineer randomly selects a batch of 12 DVD players from each day's production. The day's production is acceptable provided no more than 1 DVD player fails to meet specifications’. Otherwise, the entire day's production has to be tested.

a) What is the probability that the engineer incorrectly passes a day's production as acceptable if only 80% of the day's DVD players actually conform to speciication?
b) What is the probability that the engineer unnecessarily requires the entire day's production to be tested if in fact 90% of the DVD players conform to speciffications?

Solution

a) Let X denote the number of DVD players in the sample that fail to meet speciffications. In part (i) we want $P(X \le 1)$ with binomial parameters $n = 12$ and $\text{p} = 0.2$

$$\begin{aligned} P(X \le 1) = P(X = 0) + P(X = 1) &= {}_{12}C_0 \times (0.2)^0 (0.8)^{12} + {}_{12}C_1 \times (0.2)^1 (0.8)^{11} \\ &= 0.069 + 0.206 = 0.275 \end{aligned}$$

b) We now want $P(X > 1)$ with parameters $n = 12$ and $\text{p} = 0.1$.

$$P(X \le 1) = P(X = 0) + P(X = 1) = {}_{12}C_0 \times (0.1)^0 (0.9)^{12} + {}_{12}C_1 \times (0.1)^1 (0.9)^{11} = 0.659$$

So $\text{P(X} > 1) = 0.34$

Example 3
Bits are sent over a communications channel in packets of 12. If the probability of a bit being corrupted over this channel is 0:1 and such errors are independent, what is the probability that no more than 2 bits in a packet are corrupted?
If 6 packets are sent over the channel, what is the probability that at least one packet will contain 3 or more corrupted bits?
Let X denote the number of packets containing 3 or more corrupted bits. What is the probability that X will exceed its mean by more than 2 standard deviations?

Solution

Let C denote the number of corrupted bits in a packet. Then in the first question, we want

$$P(C \le 2) = P(C=0) + P(C=1) + P(C=2)$$

$$= {}_{12}C_0(0.1)^0(0.9)^{12} + {}_{12}C_1(0.1)^1(0.9)^{11} + {}_{12}C_2(0.1)^2(0.9)^{10}$$

$$= 0.282 + 0.377 + 0.23 = 0.889.$$

Therefore the probability of a packet containing 3 or more corrupted bits is

$$P(C \ge 3) = 1 - P(C \le 2) = 1 - 0.889 = 0.111.$$

Let X be the number of packets containing 3 or more corrupted bits. X can be modelled with a binomial distribution with parameters $n = 6$ and $\text{p} = 0.111$. The probability that at least one packet will contain 3 or more corrupted bits is:

$$P(X \ge 1) = 1 - P(X = 0) = 1 - {}_6C_0 \times (0.111)^0(0.889)^6 = 0.494.$$

The mean of X is $\text{E(X)} = 6(0.111) = 0.666$ and its standard deviation is

$$= \sqrt{6(0.111)(0.889)} = 0.77$$

So the probability that X exceeds its mean by more than 2 standard deviations is

$P(X > \mu + 2\sigma) = P(X > 2.2) = P(X \ge 3)$ since X is discrete.

Now $P(X \ge 3) = 1 - P(X \le 2) = 1 - \{P(X=0) + P(X=1) + P(X=2)\}$

$$= 1 - \left[{}_6C_0 \times (0.111)^0(0.889)^6 + {}_6C_1 \times (0.111)^1(0.889)^5 + {}_6C_2 \times (0.111)^2(0.889)^4\right]$$

$$= 1 - (0.4936 + 0.3698 + 0.1026) = 0.032$$

Exercise

1. A fair coin is tossed 10 times. What is the probability that exactly 6 heads will occur.
2. If 3% of the electric bulbs manufactured by a company are defective find the probability that in a sample of 100 bulbs exactly 5 bulbs are defective.
3. An oil exploration firm is formed with enough capital to finance 10 explorations. The probability of a particular exploration being successful is 0.1. Find mean and variance of the number of successful explorations.
4. Emily hits 60% of her free throws in basketball games. She had 25 free throws in last week's game.
   a) What is the expected number and the standard deviation of Emily's hit ?
   b) Suppose Emily had 7 free throws in yesterday's game.What is the probability that she made at least 5 hits?
5. A coin is loaded so that heads has 60% chance of showing up. This coin is tossed 3 times.
   a) What are the mean and the standard deviation of the number of heads that turned out?
   b) What is the probability that the head turns out at least twice?
   c) What is the probability that an odd number of heads turn out in 3 flips?
6. According to the 2009 current Population Survey conducted by the U.S. Census Bureau, 40% of the U.S. population 25 years old and above have completed a bachelor's degree or more. Given a random sample of 50 people 25 years old or above, what is expected number of people and the standard deviation of the number of people who have completed a bachelor's degree.
7. Joe throws a fair die six times and face number 3 appeared twice. It he incredibly lucky or unusual?
8. If the probability of being a smoker among a group of cases with lung cancer is .6, what's the probability that in a group of 8 cases you have; (a) less than 2 smokers? (b0 More than 5? (c) What are the expected value and variance of the number of smokers?
9. The manufacturer of the disk drives in one of the well-known brands of microcomputers expects 2% of the disk drives to malfunction during the microcomputer's warranty

period. Calculate the probability that in a sample of 100 disk drives, that not more than three will malfunction

10. Suppose 90% of the cars on Thika super highways does over 17 km per litre.
    a) What is the expected number and the standard deviation of cars on Thika super highways that will do over 17 km per litre.in a random sample of 15 cars ?
    b) What is the probability that in a random sample of 15 cars exactly 10 of these will do over 17 km per litre?

2.1.3 Poisson distribution

Named after the French mathematician Simeon Poisson, the distribution is used to model the number of events, (such as the number of telephone calls at a business, number of customers in waiting lines, number of defects in a given surface area, airplane arrivals, or the number of accidents at an intersection), occurring within a given time interval. Other such random events where Poisson distribution can apply includes;

- the number of hits to your web site in a day
- the number of calls that arrive in each day on your mobile phone
- the rate of job submissions in a busy computer centre per minute.
- the number of messages arriving to a computer server in any one hour.

Poisson probabilities are useful when there are a large number of independent trials with a small probability of success on a single trial and the variables occur over a period of time. It can also be used when a density of items is distributed over a given area or volume. The formula for the Poisson probability mass function is $P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$, $x=0,1,2,....$ This is abbreviated as $X \sim Po(\lambda)$. λ is the shape parameter which indicates the average number of events in the given time interval. The mean and variance of this distribution are equal ie $\mu=\sigma^2=\lambda$

Let's check to make sure that if *X* has a poisson distribution, then $\sum_{x=0}^{\infty} P(X=x)=1$. We will need to recall that $e^{\lambda}=1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+.....$ Now

$$\sum_{x=0}^{\infty} P(X=x)=\sum_{x=0}^{\infty}\frac{\lambda^x e^{-\lambda}}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{-\lambda}e^{\lambda}=e^0=1$$

Remark The major difference between Poisson and Binomial distributions is that the Poisson does not have a fixed number of trials. Instead, it uses the fixed interval of time or space in which the number of successes is recorded.

Example 1

Consider a computer system with Poisson job-arrival stream at an average of 2 per minute. Determine the probability that in any one-minute interval there will be

a) 0 jobs; b) exactly 2 jobs; c) at most 3 arrivals. d) more than 3 arrivals

Solution

Job Arrivals with λ=2

a) No job arrivals: $P(X=0)=e^{-2}=0.1353353$

b) Exactly 3 job arrivals: $P(X=3)=\frac{2^3 e^{-2}}{3!}=0.1804470$

c) At most 3 arrivals

d) $P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = \left(1 + \frac{2}{1} + \frac{2^2}{2} + \frac{2^3}{3!}\right)e^{-2} = 0.8571$

e) more than 3 arrivals $P(X > 3) = 1 - P(X \le 3) == 1 - 0.8571 = 0.1429$

Example 2

If there are 500 customers per eight-hour day in a check-out lane, what is the probability that there will be exactly 3 in line during any five-minute period?

Solution

The expected value during any one five minute period would be 500 / 96 = 5.2083333. The 96 is because there are 96 five-minute periods in eight hours. So, you expect about 5.2 customers in 5 minutes and want to know the probability of getting exactly 3.

$$P(X=3) = \frac{(-500/96)\ {}^3 e^{--500/96}}{3!} = 0.1288\,(\text{approx})$$

Example 3

If new cases of West Nile in New England are occurring at a rate of about 2 per month, then what's the probability that exactly 4 cases will occur in the next 3 months?

Solution

X ~ Poisson (λ=2/month)

$$\text{P(X = 4 in 3 months)} = \frac{(2*3)^4 e^{-(2*3)}}{4!} = \frac{6^4 e^{-(6)}}{4!} = 13.4\%$$

Exactly 6 cases?

$$\text{P(X = 6 in 3 months)} = \frac{(2*3)^6 e^{-(2*3)}}{6!} = \frac{6^6 e^{-(6)}}{6!} = 16\%$$

Exercise

1. Calculate the Poisson distribution whose λ (Average Rate of Success)) is 3 & X (Poisson Random Variable) is 6.
2. Customers arrive at a checkout counter according to a Poisson distribution at an average of 7 per hour. During a given hour, what are the probabilities that
   a) No more than 3 customers arrive?
   b) At least 2 customers arrive?
   c) Exactly 5 customers arrive?
3. Manufacturer of television set knows that on an average 5% of their product is defective. They sells television sets in consignment of 100 and guarantees that not more than 2 set will be defective. What is the probability that the TV set will fail to meet the guaranteed quality?
4. It is known from the past experience that in a certain plant there are on the average of 4 industrial accidents per month. Find the probability that in a given year will be less that 3 accidents.
5. Suppose that the change of an individual coal miner being killed in a mining accident during a year is 1.1499. Use the Poisson distribution to calculate the probability that in the mine employing 350 miners- there will be at least one accident in a year.
6. The number of road construction projects that take place at any one time in a certain city follows a Poisson distribution with a mean of 3. Find the probability that exactly five road construction projects are currently taking place in this city. (0.100819)
7. The number of road construction projects that take place at any one time in a certain city follows a Poisson distribution with a mean of 7. Find the probability that more than four road construction projects are currently taking place in the city. (0.827008)

8. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7.6. Find the probability that less than three accidents will occur next month on this stretch of road. (0.018757)
9. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 7. Find the probability of observing exactly three accidents on this stretch of road next month. (0.052129)
10. The number of traffic accidents that occur on a particular stretch of road during a month follows a Poisson distribution with a mean of 6.8. Find the probability that the next two months will both result in four accidents each occurring on this stretch of road. (0.00985)
11. Suppose the number of babies born during an 8-hour shift at a hospital's maternity wing follows a Poisson distribution with a mean of 6 an hour. Find the probability that five babies are born during a particular 1-hour period in this maternity wing. (0.160623)
12. The university policy department must write, on average, five tickets per day to keep department revenues at budgeted levels. Suppose the number of tickets written per day follows a Poisson distribution with a mean of 8.8 tickets per day. Find the probability that less than six tickets are written on a randomly selected day from this distribution. (0.128387)
13. A taxi firm has two cars which it hires out day by day. The number of demands for a car on each day is distributed as Poisson distribution with mean 1.5. Calculate the proportion of days on which neither car is used and the proportion of days on which some demands is refused
14. If calls to your cell phone are a Poisson process with a constant rate λ=0.5 calls per hour, what's the probability that, if you forget to turn your phone off in a 3 hour lecture, your phone rings during that time? How many phone calls do you expect to get during this lecture?
15. The average number of defects per wafer (defect density) is 3. The redundancy built into the design allows for up to 4 defects per wafer. What is the probability that the redundancy will not be sufficient if the defects follow a Poisson distribution?
16. The mean number of errors due to a particular bug occurring in a minute is 0.0001
    a) What is the probability that no error will occur in 20 minutes?
    b) How long would the program need to run to ensure that there will be a 99.95% chance that an error wills showup to highlight this bug?

Properties of Poisson

- The mean and variance are both equal to λ.
- The sum of independent Poisson variables is a further Poisson variable with mean equal to the sum of the individual means.
- As well as cropping up in the situations already mentioned, the Poisson distribution provides an approximation for the Binomial distribution.

2.1.4 Geometric Distribution

Suppose a Bernoulli trial with success probability p is performed repeatedly until the first success appears we want to find the probability that the first success occurs on the y^{th} trial. ie let Y denote the number of trials needed to obtain the first success. The sample space S={*s;fs;ffs, fffs ffffs …}*. This is an *infinite* sample space (though it is still discrete). What is the probability of a sample point, say $P(fffs) = P(Y = 4)$ *)*? Since successive trials are independent (this is implicit in the statement of the problem), we have

$P(fffs) = P(Y = 4) = q^3 p$ where $q = 1 - p$ and $0 \le p \le 1$

Definition: A r.v. Y is said to have a **geometric probability distribution** if and only if

$$P(Y = y) = \begin{cases} pq^{y-1} \text{for } y = 1, 2, 3..... \text{ where } q = 1 - p \\ 0 \quad\quad otherwise \end{cases}.$$

This is abbreviated as $X \sim Geo(p)$.

The only parameter for this geometric distribution is p (ie the probability of success in each trial). To be sure everything is consistent; we should check that the probabilities of all the sample points add up to 1. Now

$$\sum_{y=1}^{\infty} P(Y = y) = \sum_{y=1}^{\infty} pq^{y-1} = \frac{p}{1-q} = 1$$

Recall sum to infinity of a convergent G.P is $s = \dfrac{a}{1-r}$

The cdf of a geometric distributions is given by

$F(y) = P(Y \le y) = P(Y = 1) + P(Y = 2) + P(Y = 3) + ... + P(Y = y)$

$$= p + pq + pq^2 + ...pq^{y-1} = \frac{p(1-q^y)}{1-q} = 1 - q^y$$

Let $Y \sim$ Geo(p), then $\mu = E(Y) = \dfrac{1}{p}$ and $Var(X) = \sigma^2 = \dfrac{q}{p^2}$ Show?

Example 1

A sharpshooter normally hits the target 70% of the time.

a) Find the probability that her first hit is on the second shot
b) Find the mean and standard deviation of the number of shots required to realize the 1st hit

Solution

Let X be the random variable 'the number of shoots required to realize the 1st hit'

$x \sim Geo(0.7)$ and $P(X = x) = 0.7(1 - 0.7)^{x-1}, \ x = 1, 2, 3, ...$

a) $P(X = 2) = p(1 - \rho) = 0.7(0.3) = 0.21$

b) $\mu = \dfrac{1}{\rho} = \dfrac{1}{0.7} = 1.428571$ and $\sigma = \dfrac{\sqrt{1\text{-}p}}{p} = \dfrac{\sqrt{1-0.7}}{0.7} \approx 0.78$

Example 2

The State Department is trying to identify an individual who speaks Farsi to fill a foreign embassy position. They have determined that 4% of the applicant pool are fluent in Farsi.

a) If applicants are contacted randomly, how many individuals can they expect to interview in order to find one who is fluent in Farsi?
b) What is the probability that they will have to interview more than 25 until they find one who speaks Farsi?

Solution

a) $\mu = \dfrac{1}{\rho} = \dfrac{1}{0.04} = 25$

b) $P(X \le 25) = (1 - \rho)^n = 1 - (0.04)^{25} = 1 \ \Rightarrow \ P(X > 25) = 1 - P(X \le 25) = 0$

Example 3

From past experience it is known that 3% of accounts in a large accounting population are in error. What is the probability that 5 accounts are audited before an account in error is found? What is the probability that the first account in error occurs in the first five accounts audited?

Solution

$P(Y = 5) = 0.03(0.97)^4 = 0.02655878$ $\quad\quad$ $P(Y \le 5) = 1 - 0.97^5 = 0.14126597$

Exercise

1. Over a very long period of time, it has been noted that on Friday's 25% of the customers at the drive-in window at the bank make deposits. What is the probability that it takes 4 customers at the drive-in window before the first one makes a deposit.
2. It is estimated that 45% of people in Fast-Food restaurants order a diet drink with their lunch. Find the probability that the fourth person orders a diet drink. Also find the probability that the first diet drinker of th e day occurs before the 5th person.
3. What is the probability of rolling a sum of seven in fewer than three rolls of a pair of dice? Hint (The random variable, X, is the number of rolls before a sum of 7.)
4. In New York City at rush hour, the chance that a taxicab passes someone and is available is 15%. a) How many cabs can you expect to pass you for you to find one that is free and b) what is the probability that more than 10 cabs pass you before you find one that is free.
5. An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black ball is obtained. If we assume that each selected ball is replaced before the next one is drawn, what is;
   a) the probability that exactly n draws are needed?
   b) the probability that at least k draws are needed?
   c) the expected value and Variance of the number of balls drawn?
6. In a gambling game a player tosses a coin until a head appears. He then receives \$2*n* , where *n* is the number of tosses.
   a) What is the probability that the player receives \$8.00 in one play of the game?
   b) If the player must pay \$5.00 to play, what is the win/loss per game?
7. An oil prospector will drill a succession of holes in a given area to find a productive well. The probability of success is 0.2.
   a) What is the probability that the 3rd hole drilled is the first to yield a productive well?
   b) If the prospector can afford to drill at most 10 well, what is the probability that he will fail to find a productive well?
8. A well-travelled highway has itstraffic lights green for 82% of the time. If a person travelling the road goes through 8 traffic intersections, complete the chart to find a) the probability that the first red light occur on the nth traffic light and b) the cumulative probability that the person will hit the red light on or before the nth traffic light.
9. An oil prospector will drill a succession of holes in a given area to find a productive well. The probability of success is 0.2.
   a) What is the probability that the 3rd hole drilled is the first to yield a productive well?
   b) If the prospector can afford to drill at most 10 well, what is the probability that he will fail to find a productive well?

2.1.5 The negative binomial distribution

Suppose a Bernoulli trial is performed until the t[th] success is realized. Then the random variable "the number of trials until the t[th] success is realized" has a negative binomial distribution

Definition: A random variable X has the negative binomial distribution, also called the Pascal distribution, denoted $X \sim NB(r, p)$, if there exists an integer $n \geq 1$ and a real number $p \in (0, 1)$ such that $P(X = r + x) = {}_{r+x-1}C_x \times p^r (1-p)^x = 1, 2, 3, \ldots$

If r=1 the negative binomial distribution reduces to a geometric distribution.

2.1.6 Hyper geometric Distribution

Hyper geometric experiments occur when the trials are not independent of each other and occur due to sampling without replacement hyper-geometric probabilities involve the