Probability and Statistics (MA2001)
(Probability distributions: Bernoulli, Binomial, Poisson, Hypergeometric)

Lectures-16, 17, 18, 19

by

Dr. S. Kayal

NATIONAL INSTITUTE OF TECHNOLOGY ROURKELA

Department of Mathematics
National Institute of Technology Rourkela
Rourkela - 769008, Odisha, India

Distribution

Degenerate distribution

- The degenerate distribution characterises a distribution involving a single outcome.
- It is a deterministic distribution and takes only a single value. Examples include a two-headed coin and rolling a die whose sides all show the same number.

$$\begin{cases} P(X = a) = 1 \\ P(X \neq a) = 0 \end{cases}$$

a is a real number.

Support?

PMF?

CDF?

Mean,Median,Mode?

Variance?

Skewness,Kurtosis?

MGF?

Distributions

Bernoulli distribution (invented by Swiss mathematician Jacob Bernoulli)

- A random experiment is said to be a Bernoulli experiment if its each trial results in just two possible outcomes: success and failure.
- Each replication of a Bernoulli experiment is called a Bernoulli trial.
- A discrete random variable X with support $S_X = \{0, 1\}$ is said to follow Bernoulli distribution if its probability mass function is given by

$$f_X(x) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \\ 0, & otherwise, \end{cases} = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & otherwise, \end{cases}$$

where $0 < p < 1$.

Bernoulli distribution (cont...)

- The Bernoulli distribution represents the probability of success or failure of a single Bernoulli trial.
- The cumulative distribution function of X is given by

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1, & x \geq 1, \end{cases}$$

  where $0 < p < 1$.
- $E(X) = p$ and $Var(X) = p(1-p)$
- Moment generating function: $M_X(t) = 1 - p + e^t p$.
- Skewness: $\beta_1 = \frac{1-2p}{\sqrt{p(1-p)}}$, Kurtosis: $\gamma_1 = \frac{1-6p(1-p)}{p(1-p)}$.

Examples of Bernoulli Distribution

- Coin Tossing: When flipping a fair coin, the probability of obtaining heads (success) or tails (failure) follows a Bernoulli distribution.
- Medical Research: Researchers often categorise treatment outcomes in clinical trials as successful or unsuccessful. Bernoulli distribution aids in analysing the efficacy of a new drug or treatment, helping determine if it leads to a positive outcome.
- Market Analysis: When studying customer behaviour or market trends, variables such as purchasing decisions (buying or not buying) can be modelled using the Bernoulli distribution.
- Quality Control: Manufacturing processes often involve pass/fail situations, such as checking if a product meets certain specifications. The Bernoulli distribution can help assess the probability of success or failure in quality control, assisting in identifying and addressing potential issues.

Binomial distribution

Physical conditions for Binomial distribution: (we get the binomial distribution under the following experimental conditions)

- Each trial results in two mutually disjoint outcomes, termed as success and failure.
- The number of trials n is finite.
- The trials are independent of each other.
- The probabilty of success p is constant for each trial.

Distributions

The total number of ways of selecting r distinct combinations of N objects, irrespective of order, is

$$\binom{N}{r} = \binom{N}{N-r} = \frac{N!}{r!(N-r)!}.$$

Binomial distribution (cont...)

- A discrete type random variable X with support $S_X = \{0, 1, 2, \ldots, n\}$ is said to follow binomial distribution with parameters n (a natural number) and $p \in (0, 1)$ if it has the probability mass function

$$f_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in S_X \\ 0, & x \in S_X^c, \end{cases}$$

We denote $X \sim Binomial(n, p)$.

Binomial distribution (cont...)

- Mean=$E(X) = np$,
  $E(X^2) = n(n-1)p^2 + np$
  $Var(X) = np(1-p)$.
- Moment generating function

$$M_X(t) = (1 - p + pe^t)^n, \ t \in \mathbb{R}.$$

Skewness?

Kurtosis?

Example

Suppose you play a game that you can only either win or lose. The probability that you win any game is 55% and the probability that you lose is 45%.
(a) What is the probability that you win 15 times if you play the game 20 times?
(b) What is the probability that you lose all 20 games?
(c) What is the probability that you win all 20 games?

Solution

See the lecture!!

Distributions

Examples

1. Ten coins are thrown simultaneously. Find the probability of getting at least seven heads. (Ans: 176/1024)
2. The mean and variance of binomial distribution are 4 and 4/3, respectively. Find $P(X \geq 1)$. (Ans: 0.9986)
3. Let X be binomially distributed with parameters n and p. What is the distribution of $n - X$? (Ans: Binomial(n,1-p))
4. The moment generating function of X is $\left(\frac{2}{3} + \frac{1}{3}e^t\right)^9$. Find $P(0.2 < X < 5.6)$.
5. Let $X_1, \ldots, X_k$ be a random sample from $Binomial(3, 0.4)$. Find the distribution of $S = \sum_{i=1}^{k} X_i$.

Solution

See the lecture.

Poisson distribution (Siméon Denis Poisson)

A discrete type random variable X with support $S_X = \{0, 1, \ldots, \infty\}$ is said to follow Poisson distribution if its probability mass function is given by

$$f_X(x) = P(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x \in S_X \\ 0, & x \in S_X^c, \end{cases}$$

where $\lambda > 0$.

Some situations

- Number of occurrences in a given time interval.
- Number of accidents in a particular junction of a city.
- Number of deaths from a disease such as heart attack or due to snake bite.

Distributions

Poisson distribution from Binomial distribution

- n, the number of trials is infinitely large, that is, $n \to \infty$.
- p, the constant probability of success for each trial is infinitely small, that is, $p \to \infty$.
- $np = \lambda$ is finite, where λ is a positive number.

Under the above conditions, binomial distribution follows Poisson distribution.

Note: For the derivation of Poisson distribution from Binomial distribution, please see the lecture.

$$e = \lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x .$$

- Mean=λ=Var(X),
- Moment generating function $M_X(t) = e^{\lambda(e^t - 1)},\ t \in \mathbb{R}$.

Skewness$=\frac{1}{\sqrt{\lambda}}$.

Kurtosis$=\frac{E(X-\lambda)^4}{\lambda^2} = 3 + \frac{1}{\lambda}$.

Example

On a particular river, 5 overflow floods occur every 10 years on average. Calculate the probability of $k = 0, 1, 2, 3, 4, 5$, or 6 overflow floods in a 10-year interval, assuming the Poisson model is appropriate.

Example

It has been reported that the average number of goals in a World Cup soccer match is approximately 2.5 and the Poisson model is appropriate. Compute the probability that in a match there is exactly one goal.

Solutions

Attend the class.

Remark

Instead of average number of events, suppose we have average rate r at which the an event occurs. Then, clearly $\lambda = rt$. In this case,

$$P(\text{k events in interval t}) = \frac{e^{-rt}(rt)^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

Examples

Q1. If X is a Poisson variate with parameter λ and such that

$$P(X = 2) = 9P(X = 4) + 90P(X = 6),$$

then find mean and variance of X. (Ans:1,1)

Q2. Let X and Y be two independent Poisson variates such that $P(X = 1) = P(X = 2)$ and $P(Y = 2) = P(Y = 3)$. Find $Var(X - 2Y)$. (Ans: 14)

Q3. Consider a telephone operator who on the average handles five calls in every 3 minutes. What is the probability that there will be no calls in the next minute, at least two calls?

Q4. Consider a person who plays a series of 2500 games independently. If the probability of person winning any game is 0.002, find the probability that the person will win at least two games.

Distributions

Discrete uniform distribution

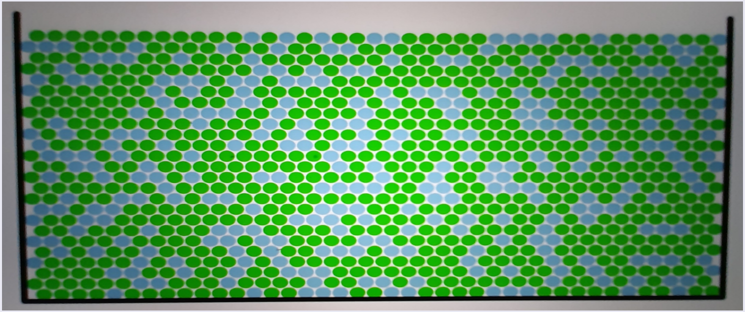
- A discrete uniform random variable X taking values $x_1 < x_2 < \ldots < x_N$ has probability mass function

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{N}, & x \in \{x_1, \ldots, x_N\} \\ 0, & otherwise, \end{cases}$$

  In particular, let $S_X = \{1, \ldots, N\}$.
- MGF $M_X(t) = \frac{e^t}{N} \frac{e^{Nt}-1}{e^t-1}$.
- Mean $E(X) = \frac{N+1}{2}$, Var(X)$=\frac{(N^2-1)}{12}$.
- Skewness=0, Kurtosis=?? (HW)

Hypergeometric distribution

An urn has 1000 balls: 700 green, 300 blue.



(a)

Sampling with replacement

- Pick one of the 1000 balls. Record color (green or blue). Put it back in the urn and shake it up. Again pick one of the 1000 balls and record color. Repeat n times.
- On each draw, the probability of green is 700/1000.
- The number of green balls drawn has a binomial distribution, with probability of success

$$p = 700/1000 = 0.7.$$

Sampling without replacement

- Pick one of the 1000 balls, record color, and set it aside. Pick one of the remaining 999 balls, record color, set it aside. Pick one of the remaining 998 balls, record color, set it aside. Repeat n times, never re-using the same ball.
- Equivalently, take n balls all at once and count them by color.
- The number of green balls drawn has a hypergeometric distribution.

Distributions

Hypergeometric distribution

- An urn contains N number of balls with K number of green balls and $N-K$ number of blue balls. A sample of n balls is drwan without replacement. What is the probability that there are k number of green balls?
- Let the random variable X be the number of green balls drawn. Then,

$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

- $E(X)=np$ and $Var(X)=\frac{np(1-p)(N-n)}{N-1}$, where $p=K/N$.

Example

An urn has 1000 balls: 700 green, 300 blue. A sample of 7 balls is drawn. What is the probability that it has 3 green balls and 4 blue balls?

- Sampling with replacement (binomial): Ans: 0.0972405
- Sampling without replacement (hypergeometric): Ans: 0.0969179

References

- *Advanced Engineering Mathematics*, Tenth Edition, by Erwin Kreyszig, John Wiley & Sons, Inc.
- *Higher Engineering Mathematics*, Sixth Edition, by B V Ramana, Tata McGraw-Hill Publisjing Company Limited.
- *An Introduction to Probability and Statistics*, Second Edition, by V. K. Rohatgi and A. K. Md. E. Saleh, Wiley.
- *Introduction to Probability and Statistics for Engineers and Scientists* by S.M. Ross, Academic Press.

Thank You