Random Variable and Probability Distributions

Dr. Upendra Pratap Singh

LNMIIT

August 27, 2023

Random Variables

Introduction

1. A **Random variable** is a function that assigns real values to each of an experiment's outcomes.
2. The probability distribution function is a mathematical function that provides the probabilities of occurrence of different possible outcomes in an experiment.

Common Probability Distributions

1. Uniform Distribution
2. Bernoulli Distribution
3. Binomial Distribution
4. Geometric Distribution
5. Gaussian Distribution
6. Poisson Distribution
7. Exponential Distribution

Uniform Distribution

Mathematical Formulation

1. A random variable X follows a uniform distribution when the probability of occurrence of an event in a given range is same, i.e.

$$P(X = i) = \frac{1}{(b - a)}$$

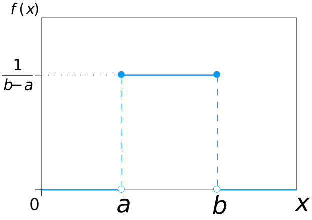

Figure: Uniform Distribution

Uniform Distribution

Statistical Measures

1. Expected value

$$\boxed{\mathbb{E}(X) = \frac{(a+b)}{2}}$$

2. Mode ?
3. Median

$$\boxed{Median = \frac{(a+b)}{2}}$$

4. Variance

$$\boxed{\sigma^2 = \frac{(b-a)^2}{12}}$$

Uniform Distribution

Applications

1. Sampling and Surveying
2. Random Number Generation
3. **Probability Modeling:** When there is no prior information suggesting one outcome to be more likely than another.

Bernoulli Distribution

Mathematical Formulation

1. A **Bernoulli** probability distribution has only *two* possible outcomes, namely 1 (success) and 0 (failure), and a *single* trial.
2. A random variable X is said to follow a **Bernoulli** distribution if its probability density function can be written as:

$$\boxed{P(X=1)=p} \tag{1}$$

$$\boxed{P(X=0)=1-p} \tag{2}$$

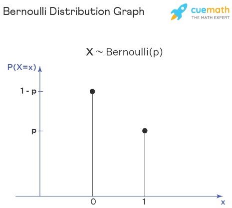

Figure: Probability distribution function of a Bernoulli Distribution

Bernoulli Distribution

Statistical Measures

1. Expected value:

$$\boxed{\mathbb{E}(X) = p}$$

2. Mode $= 1$
3. Median: with only 2 possible outcomes,

$$\boxed{Median = 1 \quad when \quad p \geq 0.5}$$

$$\boxed{Median = 0 \quad when \quad p < 0.5}$$

4. Variance

$$\boxed{\sigma^2 = p \times q}$$

Bernoulli Distribution

Key Takeaways

1. *Binary* Outcomes with only a single learnable parameter.
2. **Memorylessness:** Each trial in a Bernoulli experiment is *independent* and *memoryless*.
3. *Simplest* case of Binomial distribution characterized by $n = 1$.
4. The Bernoulli distribution serves as a *foundation* for more complex probability distributions.

Bernoulli Distribution

Applications

1. Coin Flips, inventory inspection, information theory
2. Biomedical sciences: monitoring reaction, drug effectiveness, success of a medical procedure, etc.
3. Finance, Criminal and Social justice.

Binomial Distribution

Mathematical Formulation

1. For n independent trials
2. Bernoulli: special case of Binomial with $n = 1$.
3. The binomial distribution is just n *independent* Bernoulli added up. Hence, gives the number of successes in n trials.

$$\boxed{X = Z_1 + Z_2 + Z_3 + ... + Z_n}$$

   where Z_i's are Bernoulli random variables.

4. The probability density function of a binomial distribution is given by:

$$\boxed{p(X = k) = \binom{n}{k} \, p^k \, q^{n-k}}$$

Binomial Distribution

Statistical Measures

1. Expected value:

$$\boxed{\mathbb{E}(X) = n\ p}$$

2. Mode
   - If mean μ is an integer, then mode is μ and $\mu + 1$ (provided that they have same frequency).
   - If mean μ is not an integer, then mode is the integer part of mean.
3. Median

$$\boxed{F(X \leq median) \geq 0.5}$$

4. Variance $= n\ p\ q$

Binomial Distribution

Key-Takeaways

1. Describes the number of successes in a fixed number of independent Bernoulli trials.
2. Lightweight distribution: only 2 *learnable* parameters.
3. Approximation to Gaussian Distribution: Home Work
4. **Limitation**: not suited in applications where p or n changes.

Binomial Distribution

Applications

1. **Quality Control:** inspecting a random sample for defects, predicting number of rainy days in a month, likelihood of data breaches or cyber-attacks.
2. **Genetics and Social Sciences:** studying the behavior of individuals in response to certain stimuli, analyzing the outcomes of cognitive tests.
3. **Marketing and Finance:** Analyzing the probability of default, conducting surveys to estimate the proportion of people who support a particular product.

Geometric Distribution

Mathematical Formulation

1. Models the number of trials required to achieve the first success in a sequence of *independent* Bernoulli trials.
2. Probability density function given by:

$$\boxed{P(X = n) = (1 - p)^{n-1}\, p}$$

   where the only learnable parameter p is the probability of success.
3. Discrete distribution: deals with a countable number of trials until the first success occurs.

Geometric Distribution

Statistical Measures

1. Expected value:

$$\boxed{\mathbb{E}(X) = \frac{1}{p}}$$

2. Mode $= 1$. Why ?
3. Median:

$$\boxed{Median = \lceil \frac{-1}{log_2\,(1-p)} \rceil}$$

4. Variance

$$\boxed{\sigma^2 = \frac{1-p}{p^2}}$$

Geometric Distribution

Key-Takeaways

1. **Memory-less property:** The probability of achieving the first success in the n^{th} trial is *independent* of the outcomes of previous trials.
2. *No* fixed number of trials.
3. *No* fixed upper limit.
4. **Limitation:** assumes that trials are *independent* and no upper limit on the number of trials.

Geometric Distribution

Applications

1. **Reliability and Quality Control:** modeling the number of trials until the first failure, number of claims an insurance company might receive before the first high-cost claim is mad, predicting the number of rolls of a die or spins of a roulette wheel until a specific outcome.
2. **Computer Science and IT:** modeling the number of attempts needed to successfully crack a password, the number of trials to find a specific element in a search algorithm, number of calls a customer service representative might receive before successfully resolving a customer issue.
3. **Environment and Genetics:** emergency response and rescue operations, number of attempts needed for a drug to successfully bind to a receptor site.

Gaussian Distribution

Mathematical Formulation

1. AKA normal distribution or bell curve
2. Most important and widely used probability distributions in statistics; symmetric about the mean.

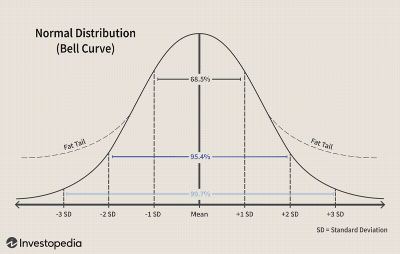

Figure: Gaussian Distribution

Gaussian Distribution

Mathematical Formulation

1. Probability density function given by:

$$\boxed{P(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, exp^{\frac{-(x-\mu)^2}{2\sigma^2}}}$$

2. 2 learnable parameters: μ and σ.

Gaussian Distribution

Statistical Measures

1. Expected value:

$$\boxed{\mathbb{E}(X) = \mu}$$

2. Mean $=$ Mode $=$ Median
3. $Variance = \sigma^2$

Gaussian Distribution

Key-Takeaways

1. Standard Normal distribution
2. **Central Limit Theorem:** The sum or average of a large number of independent random variables, regardless of their individual distribution, tends to follow a Gaussian distribution as the sample size increases.
3. **Computational Convenience:** most processes may be modelled as a Gaussian distribution

Gaussian Distribution

Applications

1. **Natural and social sciences:** modeling physical measurements, such as height, weight, temperature, and blood pressure, IQ scores, pollutant concentrations.
2. **Demography and population studies:** analyzing population characteristics, such as age distribution and income levels, mortality and birth rates.
3. **Machine learning and engineering:** modeling noise in digital images and signals, clustering (GMMs).
4. **Research:** assessing the fit of data to theoretical distributions, estimating parameters and making predictions based on the modelled distribution.

Poisson Distribution

Mathematical Formulation

A random variable X, taking on one of the values $0, 1, 2, \ldots$, is said to be a **Poisson random variable** with parameter λ, if for some $\lambda > 0$,

$$\boxed{P\{X = i\} = e^{\lambda}\frac{\lambda^i}{i!} \qquad i = 0, 1, 2, \ldots}$$

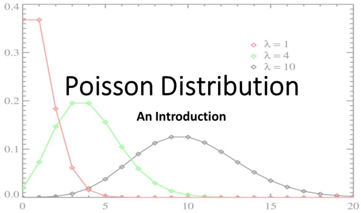

Figure: Poisson Distribution Function

Poisson Distribution

Statistical Measures

1. Expected Value
$$\boxed{\mathbb{E}(X) = \lambda}$$
2. Median remains close to the mean, particularly when λ is relatively large.
3. Mode:
$$\boxed{Mode = \lfloor \lambda \rfloor - 1}$$
4. $Variance = \lambda$

Poisson Distribution

Key Take-Aways

1. **Memorylessness**: Probability of a certain number of events occurring in a given interval of time or space does not depend on the past.
2. **Discrete**: Models events with countable integer outcomes
3. Can effectively model rare events. Why?
4. Lightweight: Mean (λ) is the only learnable parameter.
5. Can exhibit skewness. How?
6. Can approximate a binomial random variable when the binomial parameter n is large and p is small.

Poisson Distribution

Applications

To model the probability of occurrence of something in a given area, volume, measure of time, etc.

1. **Queuing Theory:** web traffic, vehicular traffic, road accidents, machine failures per unit time
2. **Inventory Management:** consumables required per unit time
3. **Social Sciences:** monitor progress, crimes per year, etc.
4. **Rare Events:** Occurrence of rare diseases, natural calamities and their spread in a time interval

Exponential Distribution

Mathematical Formulation

1. The exponential distribution is a continuous probability distribution that models the time between events in a Poisson process, where events occur continuously and independently at an average rate.
2. Probability density function is given by:

$$\boxed{P(X = x) = \lambda\, exp^{-\lambda x} \qquad x \geq 0}$$

   where λ is the only learnable parameter.
3. Defined only for $x \geq 0$ as negative time intervals are not valid.

Exponential Distribution

Statistical Measures

1. Expected value

$$\boxed{\mathbb{E}(X) = \frac{1}{\lambda}}$$

2. Variance

$$\boxed{Variance = \frac{1}{\lambda^2}}$$

Exponential Distribution

Applications

1. **Reliability and Control:** modeling the time until a component or system fails in reliability engineering, analyzing the time until an object breaks down.
2. **Environmental Studies:** modeling the time between natural events, analyzing the duration between rainfall events, modelling radioactivity phenomenon.
3. **Internet Traffic and Website Interaction:** modeling the time between user clicks, analyzing user behavior on social media platforms.
4. **Disease Modeling and Epidemiology:** modeling the time between infections or disease outbreaks in epidemiological studies.