Statistical Model

by: *Teosofi Hidayah Agung*

Discrete Distributions

- **Bernoulli Distribution:**
  The Bernoulli distribution describes a random variable that takes value 1 with probability p and value 0 with probability $1 - p$. It models a single trial of a binary experiment (e.g., success or failure).

  Ex. *Tossing a coin where heads represents success with probability $p = 0.5$.*

- **Binomial Distribution:**
  The Binomial distribution describes the number of successes in n independent Bernoulli trials, each with success probability p. It models the probability of observing a fixed number of successes in a series of trials.

  Ex. *Rolling a die 10 times and counting how many times you roll a 6 (with probability $p = 1/6$ for rolling a 6).*

- **Negative Binomial Distribution:**
  The Negative Binomial distribution models the number of failures before achieving r successes in a sequence of independent Bernoulli trials with probability p of success. It is a generalization of the geometric distribution.

  Ex. *Tossing a coin and counting how many tails occur before getting 5 heads.*

- **Geometric Distribution:**
  The Geometric distribution models the number of trials needed to achieve the first success in a series of independent Bernoulli trials with success probability p.

  Ex. *Rolling a die until you get the number 3, where success is defined as rolling a 3.*

- **Hypergeometric Distribution:**
  The Hypergeometric distribution models the number of successes in n draws from a finite population of N objects, without replacement, where M of the objects are classified as successes.

  Ex. *Drawing 5 cards from a deck of 52 cards without replacement, and counting how many are aces.*

- **Multinomial Distribution:**
  The Multinomial distribution is a generalization of the binomial distribution, describing the probabilities of counts for k different possible outcomes in n independent trials, with each outcome having a probability p_i.

  Ex. *Rolling a die 10 times and recording the number of times each face (1 through 6) appears.*

- **Poisson Distribution:**
  The Poisson distribution models the number of events that occur in a fixed interval of time or space, assuming events occur independently and at a constant average rate μ.

  Ex. *Counting the number of emails you receive in an hour, assuming the average rate is 5 emails per hour.*

- **Discrete Uniform Distribution:**
  The Discrete Uniform distribution models a random variable that takes integer values within a range $[a, b]$ with equal probability. It is often used to model situations where all outcomes are equally likely.

  Ex. *Choosing a random number between 1 and 10, where each number has an equal chance of being selected.*

Continuous Distributions

- **Uniform Distribution:**
  The Uniform distribution models a random variable that is equally likely to take any value between two limits a and b. It has a constant probability density over the interval.

  Ex. *Selecting a random time between 2 PM and 4 PM for a meeting.*

- **Normal Distribution:**
  The Normal distribution (or Gaussian distribution) is one of the most important distributions in statistics. It models a continuous random variable with a symmetric, bell-shaped probability density centered at μ, with spread determined by σ^2.

  Ex. *The heights of adults in a population tend to follow a normal distribution with a mean height and standard deviation.*

- **Gamma Distribution:**
  The Gamma distribution is used to model waiting times in processes where events occur continuously and independently at a constant average rate. It is parameterized by a shape parameter κ and a scale parameter θ.

  Ex. *The time until the 10th call is received at a call center, where calls arrive at a constant rate.*

- **Exponential Distribution:**
  The Exponential distribution is a special case of the Gamma distribution, modeling the time between events in a Poisson process. It has a single parameter θ, which represents the mean waiting time between events.

  Ex. *The time between consecutive arrivals at a bus stop where buses arrive randomly at an average rate.*

- **Weibull Distribution:**
  The Weibull distribution is often used in reliability analysis and failure time modeling. It is flexible, with two parameters: θ, the scale, and β, the shape parameter. The shape parameter controls the failure rate behavior over time.

  Ex. *The lifespan of a machine part where early failures are more likely.*

- **Pareto Distribution:**
  The Pareto distribution models heavy-tailed phenomena, where a few large events dominate the distribution (e.g., wealth distribution). It is parameterized by a shape parameter κ and a scale parameter θ.

  Ex. *Modeling the distribution of wealth in a population, where a small number of people hold most of the wealth.*

- **Beta Distribution:**
  The Beta distribution models random variables that take values in the interval $[0, 1]$, and it is commonly used to model proportions or probabilities. It is parameterized by two shape parameters α and β.

  Ex. *The probability that a website visitor will click on an ad, based on past data.*

- **Cauchy Distribution:**
  The Cauchy distribution has a bell shape similar to the normal distribution but with heavier tails. It is used in situations where data contains extreme outliers. The mean and variance do not exist for this distribution.

  Ex. *Modeling the distribution of data with extreme outliers, such as returns on certain financial investments.*

- **Chi-Squared Distribution:**
  The Chi-Squared distribution is the distribution of a sum of squared independent standard normal variables. It is widely used in hypothesis testing, particularly in the context of variance estimation and goodness-of-fit tests. It is parameterized by v, the degrees of freedom.

  Ex. *Testing whether observed frequencies in a categorical dataset match the expected frequencies (chi-squared goodness-of-fit test).*

Distribution Table

Discrete Distributions

Name Distribution	Notation and Parameter	Discrete PDF $f(x)$	Expectation $E(X)$	Variance $\mathrm{Var}(X)$	MGF $M_X(t)$
Bernoulli	$X \sim B(1,p)$	$p^x(1-p)^{1-x}$	p	$p(1-p)$	$1-p+pe^t$
Binomial	$X \sim B(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	np	$np(1-p)$	$(1-p+pe^t)^n$
Negative Binomial	$X \sim NB(r,p)$	$\binom{x-1}{r-1}p^r(1-p)^x$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$	$\left(\dfrac{p}{1-(1-p)e^t}\right)^r$
Geometric	$X \sim G(p)$	$p(1-p)^{x-1}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{p}{1-(1-p)e^t}$
Hypergeometric	$X \sim H(n,M,N)$	$\dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$n\dfrac{M}{N}$	$n\dfrac{M}{N}\left(1-\dfrac{M}{N}\right)\dfrac{N-n}{N-1}$	*
Multinomial	$X \sim M(n,p_1,\ldots,p_k)$	$\dfrac{n!}{x_1!\cdots x_k!}p_1^{x_1}\cdots p_k^{x_k}$	np_i	$np_i(1-p_i)$	$\left(\sum_{i=1}^{k}p_ie^{t_i}\right)^n$
Poisson	$X \sim P(\mu)$	$\dfrac{e^{-\mu}\mu^x}{x!}$	μ	μ	$e^{\mu(e^t-1)}$
Discrete Uniform	$X \sim U(a,b)$	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{tb}-e^{ta}}{t(b-a)}$

Continuous Distributions

Name Distribution	Notation and Parameter	Continuous PDF $f(x)$	Expectation $E(X)$	Variance $\mathrm{Var}(X)$	MGF $M_X(t)$
Uniform	$X \sim UNIF(a,b)$	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{bt}-e^{at}}{(b-a)t}$
Normal	$X \sim N(\mu,\sigma^2)$	$\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t+\frac{\sigma^2 t^2}{2}}$
Gamma	$X \sim GAM(\theta,\kappa)$	$\dfrac{1}{\theta^\kappa\Gamma(\kappa)}x^{\kappa-1}e^{-x/\theta}$	$\kappa\theta$	$\kappa\theta^2$	$\left(\dfrac{1}{1-\theta t}\right)^\kappa$
Exponential	$X \sim EXP(\theta)$	$\dfrac{1}{\theta}e^{-x/\theta}$	θ	θ^2	$\dfrac{1}{1-\theta t}$
Weibull	$X \sim WEI(\theta,\beta)$	$\dfrac{\beta}{\theta^\beta}x^{\beta-1}e^{-(x/\theta)^\beta}$	$\theta\Gamma\left(1+\dfrac{1}{\beta}\right)$	$\theta^2\left[\Gamma\left(1+\dfrac{2}{\beta}\right)-\Gamma^2\left(1+\dfrac{1}{\beta}\right)\right]$	*
Pareto	$X \sim PAR(\theta,\kappa)$	$\dfrac{\kappa}{\theta(1+x/\theta)^{\kappa+1}}$	$\dfrac{\theta}{\kappa-1}$	$\dfrac{\theta^2\kappa}{(\kappa-1)^2(\kappa-2)}$	**
Beta	$X \sim BETA(\alpha,\beta)$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$	$\dfrac{\alpha}{\alpha+\beta}$	$\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	-
Cauchy	$X \sim CAU(\eta,\theta)$	$\dfrac{1}{\pi\eta\left[1+((x-\theta)/\eta)^2\right]}$	**	**	**
Chi-Squared	$X \sim \chi^2(v)$	$\dfrac{1}{2^{v/2}\Gamma(v/2)}x^{v/2-1}e^{-x/2}$	v	$2v$	$\left(\dfrac{1}{1-2t}\right)^{v/2}$

*Not tractable.
**Does not exist.