Common Probability Distributions

1. **Distribution**: A description or function that shows the possible values of a variable and how often they occur.
2. **Probability**: A measure of the likelihood that a specific event will occur, expressed as a number between 0 and 1.

Discrete Probability Distributions

Discrete: Refers to data or variables that can only take specific, separate values (e.g., integers), with no intermediate values between them.

Discrete Probability Distribution: A distribution that gives the probabilities of occurrence of different values for a discrete random variable.

Bernoulli Distribution

The Bernoulli distribution is a discrete probability distribution for a random variable which has only two possible outcomes: 0 (failure) and 1 (success).The probability of success is denoted by (p), and the probability of failure is (1 - p).

```r
# Using dbern() from Rlab:
library(Rlab)
```

```
## Rlab 4.0 attached.
```

```
##
## Attaching package: 'Rlab'
```

```
## The following objects are masked from 'package:stats':
##
##     dexp, dgamma, dweibull, pexp, pgamma, pweibull, qexp, qgamma,
##     qweibull, rexp, rgamma, rweibull
```

```
## The following object is masked from 'package:datasets':
##
##     precip
```

```r
# Calculate Bernoulli probabilities
dbern(c(0, 1), prob = 0.7)
```

```
## [1] 0.3 0.7
```

Binomial Distribution

The Binomial distribution generalizes the Bernoulli distribution to multiple trials. It describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success (p).The Binomial distribution is defined by two parameters: (n) (number of trials) and (p) (probability of success).

```
# dbinom(x, size, prob): Computes the probability mass function (PMF).
# Probability of getting exactly 3 successes in 10 trials with success of 0.5.
dbinom(3, size = 10, prob = 0.5)
```

```
## [1] 0.1171875
```

```
# pbinom(q, size, prob): Computes the cumulative distribution function (CDF).
# Probability of getting 3 or fewer successes in 10 trials with success of 0.5.
pbinom(3, size = 10, prob = 0.5)
```

```
## [1] 0.171875
```

```
# qbinom(p, size, prob):Computes the quantile function.Number of successes
# corresponds to the 0.5 quantile in 10 trials with success probability 0.5.
qbinom(0.5, size = 10, prob = 0.5)
```

```
## [1] 5
```

```
# rbinom(n, size, prob): Generates random samples.
# Generate 5 random samples from a Binomial distribution with 10 trials and
# success probability 0.5.
rbinom(5, size = 10, prob = 0.5)
```

```
## [1] 7 7 6 4 8
```

Poisson Distribution

The Poisson distribution models the number of events occurring within a fixed interval of time or space, given the average number of times the event occurs over that interval. It is defined by the parameter(lambda), which is the average rate of occurrence.

```
# dpois(x, lambda):Computes the probability mass function (PMF).
# Probability of observing exactly 3 events when the average rate is 2.
dpois(3, lambda = 2)
```

```
## [1] 0.180447
```

```
# ppois(q, lambda):Computes the cumulative distribution function (CDF).
#Probability of observing 3 or fewer events when the average rate is 2.
ppois(3, lambda = 2)
```

```
## [1] 0.8571235
```

```r
# qpois(p,lambda):Computes the quantile function.
#Number of events corresponding to the 0.5 quantile when the average rate is 2.
qpois(0.5, lambda = 2)
```

```
## [1] 2
```

```r
# rpois(n, lambda):Generates random samples.
# Generate 5 random samples from  a Poisson dist. with an average rate of 2.
rpois(5, lambda = 2)
```

```
## [1] 0 4 8 2 2
```

Geometric Distribution

The Geometric distribution models the number of trials needed to get the first success in a series of independent and identically distributed Bernoulli trials. It is defined by the parameter (p), the probability of success on each trial.

```r
# dgeom(x, prob):Computes the probability mass function (PMF).
# Probability needed exactly 3 trial to get first success with success of 0.5.
dgeom(3, prob = 0.5)
```

```
## [1] 0.0625
```

```r
# pgeom(q, prob):Computes the cumulative distribution function (CDF).
# Probability needed 3 or fewer trial to get first success with success of 0.5.
pgeom(3, prob = 0.5)
```

```
## [1] 0.9375
```

```r
# qgeom(p,prob):Computes the quantile function.
# Number of trials corresponds to the 0.5 quantile with success probability 0.5.
qgeom(0.5, prob = 0.5)
```

```
## [1] 0
```

```r
# rgeom(n, prob):Generates random samples.
# Generate 5 random samples from a Geometric distribution with a success of 0.5.
rgeom(5, prob = 0.5)
```

```
## [1] 0 0 3 0 2
```

Hypergeometric Distribution

The Hypergeometric distribution models the number of successes in a sample of size (n) drawn without replacement from a population of size (N) that contains (K) successes.It is defined by the parameters (N) (population size) ,(K) (number of successes in the population), and (n) (sample size).

```r
# dhyper(x, m, n, k): Computes the probability mass function (PMF).
dhyper(3, m = 20, n = 30, k = 10)
```

```
## [1] 0.2259296
```

```r
# phyper(q, m, n, k):Computes the cumulative distribution function (CDF).
phyper(3, m = 20, n = 30, k = 10)
```

```
## [1] 0.3649683
```

```r
# qhyper(p,m,n,k):Computes the quantile function.
qhyper(0.5, m = 20, n = 30, k = 10)
```

```
## [1] 4
```

```r
# rhyper(nn, m, n, k):Generates random samples.
rhyper(5, m = 20, n = 30, k = 10)
```

```
## [1] 6 4 5 5 6
```

Multinomial Distribution

The Multinomial distribution generalizes the Binomial distribution to more than two outcomes.It models the counts of each outcome in a fixed number of independent trials, where each trial results in one of (k) possible outcomes each with its own probability.

```r
# rmultinom(n, size, prob): Generates random samples.
# Generate 5 random samples from a Multinomial distribution with 10 trials and
# probabilities 0.2, 0.3, and 0.5
rmultinom(5, size = 10, prob = c(0.2, 0.3, 0.5))
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    2    0    2    2    2
## [2,]    3    4    1    4    4
## [3,]    5    6    7    4    4
```

```r
# dmultinom(x, size, prob): Computes the probability mass function (PMF).
# Probability of getting 2,3 and 5 outcomes in 10 trials with probability   0.2, # 0.3, and 0.5.
dmultinom(c(2, 3, 5), size = 10, prob = c(0.2, 0.3, 0.5))
```

```
## [1] 0.08505
```

Negative Binomial Distribution

The Negative Binomial distribution models the number of trials needed to get (r) successes in a series of independent and identically distributed Bernoulli trials. It is defined by two parameters: (r) (number of successes) and (p) (probability of success on each trial).

```r
# dnbinom(x, size, prob): Computes the probability mass function (PMF).
# Probability needed exactly 3 trial to get 2 success with a success probability # of 0.5.
dnbinom(3, size = 2, prob = 0.5)
```

```
## [1] 0.125
```

```r
# pnbinom(q, size, prob): Computes the cumulative distribution function (CDF).
# Probability of needing 3 or fewer trials to get 2 successes with a success
# probability of 0.5.
pnbinom(3, size = 2, prob = 0.5)
```

```
## [1] 0.8125
```

```r
# qnbinom(p, size, prob): Computes the quantile function.
qnbinom(0.5, size = 2, prob = 0.5)
```

```
## [1] 1
```

```r
# rnbinom(n, size, prob): Generates random samples.
rnbinom(5, size = 2, prob = 0.5)
```

```
## [1] 1 2 2 4 2
```

Continouos probability distributions

Continuous: Refers to data or variables that can take any value within a given range, including fractions and decimals.

Continuous Probability Distribution: A distribution that gives the probabil -ities of occurrence of different values for a continuous random variable.

1. Density Function: dunif(x, min, max)
2. Cumulative Distribution Function: punif(x, min, max)
3. Quantile Function: qunif(p, min, max)
4. Random Generation: runif(n, min, max)

Uniform Distribution

The uniform distribution is a continuous probability distribution where all intervals of the same length within the distribution's range are equally probable.parameter of the distribution is a and b and distribution ranges from a to b.

```r
# Density
dunif(0.5, min=0, max=1)
```

```
## [1] 1
```

```r
# Cumulative Probability
punif(0.5, min=0, max=1)
```

```
## [1] 0.5
```

```r
# Quantile
qunif(0.5, min=0, max=1)
```

```
## [1] 0.5
```

```r
# Random Generation
runif(10, min=0, max=1)
```

```
##  [1] 0.366495374 0.250636494 0.453805685 0.001326805 0.456806588 0.018293457
##  [7] 0.436497648 0.819722436 0.194704476 0.307841394
```

Normal Distribution

The normal distribution, also known as the Gaussian distribution,is a continuous probability distribution characterized by its bell-shaped curve.Its parameters are the mean (mu) and standard deviation (sigma).The distribution ranges from negative infinity to positive infinity.

```r
# Density
dnorm(0, mean=0, sd=1)
```

```
## [1] 0.3989423
```

```r
# Cumulative Probability
pnorm(0, mean=0, sd=1)
```

```
## [1] 0.5
```

```r
# Quantile
qnorm(0.5, mean=0, sd=1)
```

```
## [1] 0
```

```r
# Random Generation
rnorm(10, mean=0, sd=1)
```

```
##  [1] -1.60495052  0.59692099  0.33849944 -0.28642175  0.90937061 -0.02687466
##  [7]  1.13553678 -0.37965655  0.17506746 -0.45561797
```

Exponential Distribution

The exponential distribution is used to model the time between events in a Poisson process.The parameter of distribution is rate(lambda) which is the average number of events per unit of time.The range of distribution is from 0 to infinity.

```r
# Density
dexp(1, rate=1)
```

```
## [1] 0.3678794
```

```r
# Cumulative Probability
pexp(1, rate=1)
```

```
## [1] 0.6321206
```

```r
# Quantile
qexp(0.5, rate=1)
```

```
## [1] 0.6931472
```

```r
# Random Generation
rexp(10, rate=1)
```

```
##  [1] 1.12571171 2.26827758 0.67619858 0.06000073 1.59896198 0.12484201
##  [7] 0.39613917 0.91423630 0.09346710 1.05611241
```

Chi-Square Distribution

The chi-square distribution is used in hypothesis testing and construction of confidence intervals.The parameter of distribution is degrees of freedom(df).The range of distribution is from 0 to infinity.

```r
# Density
dchisq(2, df=3)
```

```
## [1] 0.2075537
```

```r
# Cumulative Probability
pchisq(2, df=3)
```

```
## [1] 0.4275933
```

```r
# Quantile
qchisq(0.95, df=3)
```

```
## [1] 7.814728
```

```r
# Random Generation
rchisq(10, df=3)
```

```
##  [1] 3.7003621 1.8301975 0.8574080 1.0636134 1.6261234 9.5074948 3.2339218
##  [8] 0.4815075 0.4022714 3.3988025
```

Beta Distribution

The beta distribution is a continuous probability distribution defined on the interval [0, 1].The parameter of distribution A larger   skews the distribution toward 1. A larger   skews the distribution toward 0.

```
# Density
dbeta(0.5, shape1=2, shape2=5)
```

```
## [1] 0.9375
```

```
# Cumulative Probability
pbeta(0.5, shape1=2, shape2=5)
```

```
## [1] 0.890625
```

```
# Quantile
qbeta(0.5, shape1=2, shape2=5)
```

```
## [1] 0.26445
```

```
# Random Generation
rbeta(10, shape1=2, shape2=5)
```

```
##  [1] 0.25960130 0.15786317 0.57769134 0.08722195 0.20037433 0.21270845
##  [7] 0.11211120 0.28335292 0.06533800 0.08056235
```

Gamma Distribution

The gamma distribution is used to model waiting times.The parameter of the distribution is shape(alpha) and rate(beta).The range of distribution is from 0 to infinity.

```
# Density
dgamma(2, shape=2, rate=1)
```

```
## [1] 0.2706706
```

```
# Cumulative Probability
pgamma(2, shape=2, rate=1)
```

```
## [1] 0.5939942
```

```
# Quantile
qgamma(0.5, shape=2, rate=1)
```

```
## [1] 1.678347
```

```r
# Random Generation
rgamma(10, shape=2, rate=1)
```

```
##  [1] 1.9246876 1.0185482 0.8988618 3.9076253 2.5543903 0.4422436 1.1237166
##  [8] 0.7506261 3.7287815 1.1979322
```

F Distribution

The F distribution is used primarily in ANOVA and regression analysis.The Parameter of distribution is degrees of freedom(df1) and degrees of freedom(df2) .The range of distribution is from 0 to infinity.

```r
# Density
df(2, df1=5, df2=2)
```

```
## [1] 0.1320704
```

```r
# Cumulative Probability
pf(2, df1=5, df2=2)
```

```
## [1] 0.6339381
```

```r
# Quantile
qf(0.95, df1=5, df2=2)
```

```
## [1] 19.29641
```

```r
# Random Generation
rf(10, df1=5, df2=2)
```

```
##  [1] 1.17613642 0.06658623 0.99661559 0.85106655 1.14180093 4.54976234
##  [7] 1.28353168 2.57181692 1.71995517 0.97021687
```

t Distribution

The t distribution is used in hypothesis testing and confidence intervals for small sample sizes.The parameter of distribution is degrees of freedom(df).The range of distribution is from negative infinity to positive infinity.

```r
# Density
dt(0, df=10)
```

```
## [1] 0.3891084
```

```r
# Cumulative Probability
pt(0, df=10)
```

```
## [1] 0.5
```

```r
# Quantile
qt(0.95, df=10)
```

```
## [1] 1.812461
```

```r
# Random Generation
rt(10, df=10)
```

```
##  [1]  0.708447591  0.126012518 -0.443491132  0.828032864  0.143699192
##  [6]  0.243641529  0.692661506  0.003804288 -0.194682084  0.266994939
```

Weibull Distribution

The Weibull distribution is used to model the time until an event occurs.The parameter of distribution is shape(k) and scale(lambda).The range of distribut- ion is from 0 to infinity.

```r
# Density
dweibull(2, shape=2, scale=1)
```

```
## [1] 0.07326256
```

```r
# Cumulative Probability
pweibull(2, shape=2, scale=1)
```

```
## [1] 0.9816844
```

```r
# Quantile
qweibull(0.5, shape=2, scale=1)
```

```
## [1] 0.8325546
```

```r
# Random Generation
rweibull(10, shape=2, scale=1)
```

```
##  [1] 0.9404379 0.9152061 0.3002711 1.8752539 1.1803144 0.1949046 1.9008095
##  [8] 1.1675233 0.7016604 1.1851404
```

Log-Normal Distribution

The log-normal distribution is used to model data that is skewed to the right. The parameter of distribution is mean log(mu) and sdlog(sigma).The range of distribution is from 0 to infinity.

```r
# Density
dlnorm(2, meanlog=0, sdlog=1)
```

```
## [1] 0.156874
```

```r
# Cumulative Probability
plnorm(2, meanlog=0, sdlog=1)
```

```
## [1] 0.7558914
```

```r
# Quantile
qlnorm(0.5, meanlog=0, sdlog=1)
```

```
## [1] 1
```

```r
# Random Generation
rlnorm(10, meanlog=0, sdlog=1)
```

```
##  [1] 0.9062514 0.5888558 2.4631290 9.7273292 1.9772515 1.1479336 0.4549555
##  [8] 0.8291270 4.5547893 1.7111925
```

Cauchy Distribution

The Cauchy distribution is used to model data with heavy tails.The parameter of distribution is location and scale.The range of distribution is from negative infinity to positive infinity.

```r
# Density
dcauchy(0, location=0, scale=1)
```

```
## [1] 0.3183099
```

```r
# Cumulative Probability
pcauchy(0, location=0, scale=1)
```

```
## [1] 0.5
```

```r
# Quantile
qcauchy(0.5, location=0, scale=1)
```

```
## [1] 0
```

```r
# Random Generation
rcauchy(10, location=0, scale=1)
```

```
##  [1]  1.6976950 -1.4004554 -1.4815424 -1.3055777 -0.3142037  5.0169382
##  [7] -1.3608248 -0.3108277 -5.4197604 -6.0078810
```

Logistic Distribution

The logistic distribution is used to model growth processes.The parameter of distribution is location and scale.The range of distribution is from negative infinity to positive infinity.

```r
# Density
dlogis(0, location=0, scale=1)
```

```
## [1] 0.25
```

```r
# Cumulative Probability
plogis(0, location=0, scale=1)
```

```
## [1] 0.5
```

```r
# Quantile
qlogis(0.5, location=0, scale=1)
```

```
## [1] 0
```

```r
# Random Generation
rlogis(10, location=0, scale=1)
```

```
##  [1] -0.8434506 -0.4184938  1.2048981  1.7283547 -1.9622324 -1.9279192
##  [7]  1.9380663  1.9112864  1.5417776  2.2913817
```