

CS544 Module4

Suresh Kalathur

Module4

- Data Distributions
 - Discrete
 - Continuous
- Random variables
 - # of students showing up for the class
 - Income of a person, height of a player, etc.

Discrete Distributions

- Discrete Random Variables
 - Support
 - All possible values for the random variable
 - Probability Mass Function (PMF)
 - $f_X(x)$ *i. e.* $P(X = x)$
 - Mean or Expected Value, variance, standard deviation
 - Cumulative Distribution Function (CDF)
 - $F_X(x)$ *i. e.* $P(X \leq x)$

Bernoulli Trials

- Binomial coefficients

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

- Bernoulli Trials
 - Random experiment with two possible outcomes
 - Probability of success, p
 - Probability of failure, $1-p$
 - Review PMF, mean, and variance
 - Repeated trials
 - The trials are independent,
 - Each trial has two possible outcomes (success and failure)
 - The probability of success remains the same from trial to trial.

Binomial Distribution

- Probability distribution for
 - the number of successes in a sequence of Bernoulli trials.
- Two parameters
 - n , the number of trials
 - p , the probability of success
- Review PMF, mean, and variance
- R – `dbinom`, `pbinom`, `qbinom`, `rbinom`

The 4 Functions

- **d**<name>(x, ...) $f_X(x)$ i. e. $P(X = x)$
 - Probability Density function
- **p**<name>(x, ...) $F_X(x)$ i. e. $P(X \leq x)$
 - Cumulative Distribution function
- **q**<name>(p, ...) smallest x such that $F_X(x) \geq p$
 - Quantile function
- **r**<name>(n, ...)
 - **n** random values from the distribution

Hypergeometric Distribution

- Outcomes dependent on previous outcomes
- Sample data selected without replacement
- Three parameters
 - M , # of events of interest
 - N , # of events not of interest
 - K , the sample size without replacement
- Review PMF, mean, and variance
- R – `dhyper`, `phyper`, `qhyper`, `rhyper`

Geometric Distribution

- # of failures before a success in a sequence of Bernoulli trials
- One parameter
 - p , probability of success
- Review PMF, mean, and variance
- R – `dgeom`, `pgeom`, `qgeom`, `rgeom`

Negative Binomial Distribution

- # of failures until a total or “ r ” successes in a sequence of Bernoulli trials
- Two parameters
 - p , probability of success
 - r , the total number of successes
- Review PMF, mean, and variance
- R – `dnbinom`, `pnbinom`, `qnbino`, `rnbinom`

Poisson Distribution

- Model the frequency with which a specified event occurs during a particular period of time
- One parameter
 - λ , average number of events per unit of time [0,1]
- Review PMF, mean, and variance
- R – dpois, ppois, qpois, rpois

Discrete Uniform Distribution

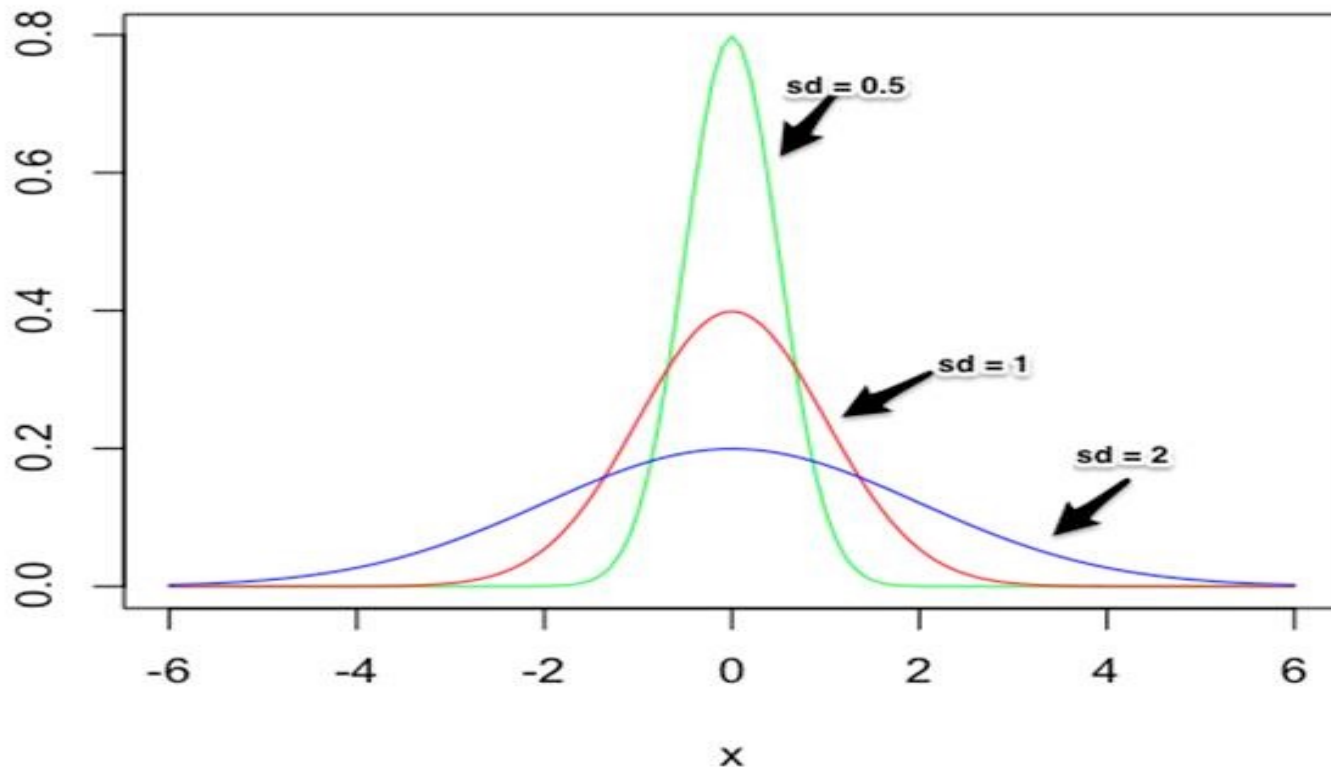
- Each value is equally likely
- Review PMF, CDF, mean, and variance
- R – dunif, punif, qunif, runif
- sample() function
 - sample(x, size, replace = FALSE, prob = NULL)

Continuous Distributions

- Continuous uniform distribution
 - Two parameters (min and max)
 - Review PDF and CDF
- R functions
 - `dunif`, `punif`, `qunif`, `runif`

Normal Distribution

- Determined by the mean (μ) and standard deviation (σ)
- R functions (dnorm, pnorm, qnorm, rnorm)



Exponential Distribution

- Waiting times, time between arrivals, etc.
- One parameter
 - λ , mean number of arrivals per unit of time
- R functions
 - `dexp`, `pexp`, `qexp`, `rexp`

Term Project

- Review Project Requirements
- Review Sample Past Projects