Probability Distributions

Dr. Paige Rutner

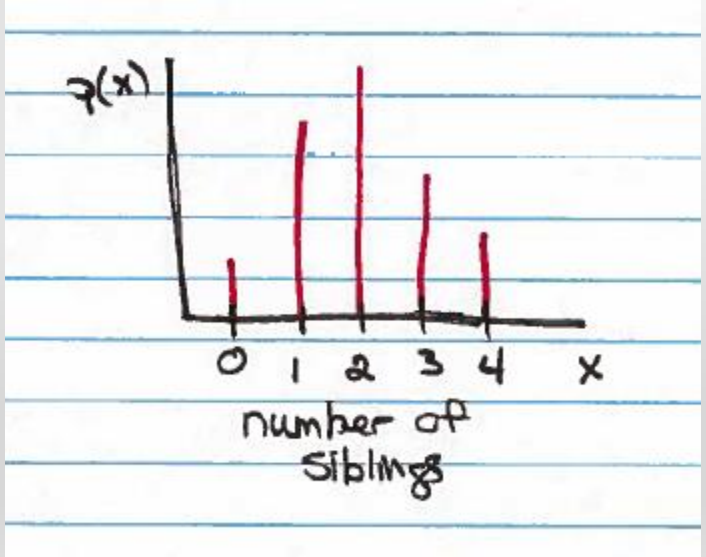
Associate Professor of Practice

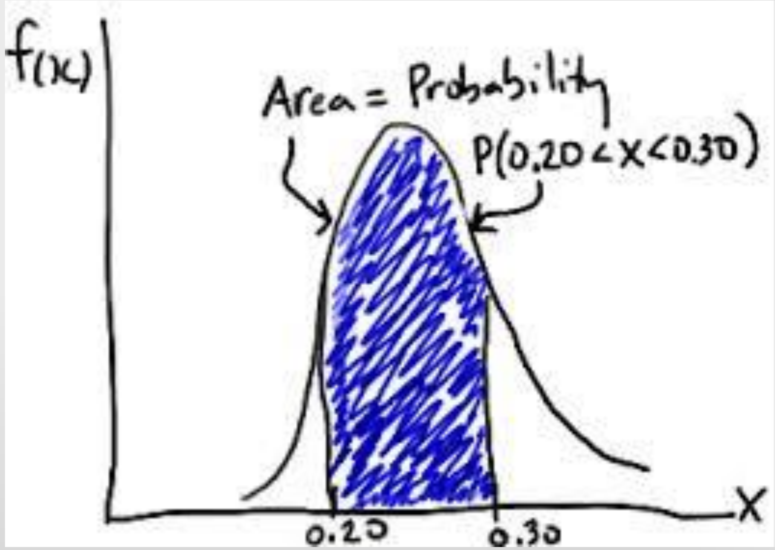
Information Systems and Quantitative Sciences

Selected Topics from Westfall & Henning, Ch 2 - 4

Probability Distribution Function

- Shows
  - the set of possible values a random variable can take, and
  - the likelihood that those values will occur.





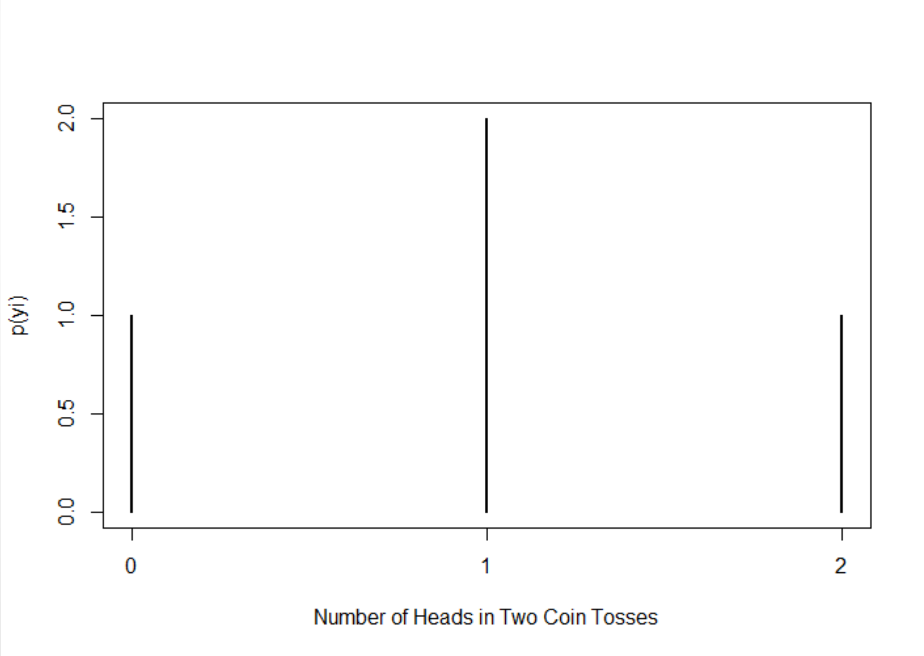
Discrete and Continuous Variables

- All observed data are actually discrete
- However, we will frequently treat those data as though they were continuous
- Why?
  - Statistics computed from discrete data are often continuous
  - A continuous model may be a “good” model even for discrete data
  - DATA* produced by a continuous model can be categorized into discrete groups

Discrete pdf

- Sometimes called a *probability mass function*
- Requirements
  - $p(y) \geq 0, for\ y \in S$
  - $\sum_{y \in S} p(y) = 1$

Number of heads in two coin tosses (y_i)	$p(y_i)$
0	.25
1	.50
2	.25



Multinomial distribution

- Shows number of successes from n independent trials with a discrete number of possible outcomes.
- Each outcome has a constant probability of success across trials.
- For example, drawing an even card, odd card, or face card from a full deck.
- $p(y|\theta) = \pi_1^{I(y=Even)}\pi_2^{I(y=Odd)}\pi_3^{I(y=Face)}, for\ y \in \{Even, Odd, Face\}$

  where θ = (π_1, π_2, π_3), and

  I(condition) is 1 if the condition is true, and 0 otherwise.

Outcome (y)	p(y|θ)
Even (2,4,6,8,10)	0.3846
Odd (A,3,5,7,9)	0.3846
Face (J,Q,K)	0.2308

Bernoulli Distribution

- Discrete pdf with two possible outcomes
- π defined as probability of "success"
- $p(y|\pi) = \begin{cases} 1-\pi, & for\ y=0 \\ \pi, & for\ y=1 \end{cases}$
- $p(y|\pi) = \pi^{y}(1-\pi)^{1-y},\ for\ y \in \{0,1\}$
- Basis for other discrete pdfs
  - Binomial – number of successes in *n* trials
  - Geometric – number of failures before first success from *n* trials
  - Negative Binomial – number of failures before the x^{th} success from *n* trials

Poisson Distribution

- Describes the number of independent events that occur in some area of opportunity
- λ is the theoretical average number of events per area of opportunity
- $p(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, for\ y = 0, 1, 2, \ldots$

  where e = 2.71828…

Discrete pdf probabilities

- Probability of an Event A for a Discrete RV

$$\Pr(Y \in A) = \sum_{y \in A} p(y)$$

- The probability that the discrete variable Y is in the set A is equal to the sum of the probabilities of the individual outcomes that are in the set A

- For example, suppose that the distribution of fatal auto accidents per day in a city follows a Poisson distribution with $\lambda = .3$.
  - What is the probability of having 0 accidents during a day?
  - What is the probability of having 1 accident during a day?
  - What is the probability of having 1 or more accidents during a day?

Probability simulation

- Data from surveys are often recorded on a scale with levels from 1 to 5. Here are some responses from n=10 people surveyed by an automobile retailer about their level of satisfaction:

  3, 4, 3, 5, 5, 5, 5, 5, 5, 4

- Create a summary table and a needle plot of the data
- Using list form, show a discrete pdf that could have produced these data. It is possible to have dissatisfied customers (satisfaction of 1 or 2), so none of the probabilities should be 0.
- Use R to simulate n=10 values from this distribution
- Is this a good model?

Some R functions for discrete pdfs

- Random number generators
  - rpois()
  - rbinom()
  - sample()
- Summarizing data
  - table()
- Visualizing data
  - barplot()
  - plot()

Continuous pdf

- Sometimes called *probability density function*
- Given that the random variable *y* is a function of x.

$$y = f(x)$$

- Requirements for continuous pdf:
  - The value of *y* is greater than or equal to 0 for all values of *x.*

$$p(y) \geq 0, for\ all\ y \in S.$$

  - The total area under the curve of the function is equal to 1.

$$\int_{y \in S} p(y)\, dy = 1.$$

Continuous pdf probability

- Probability of an Event A for a Continuous RV

$$\Pr(Y \in A) = \int_{y \in A} p(y)dy$$

- Expressed another way

$$P(c < Y < d) = \int_{c}^{d} f(y)dy$$

Example problem

- Given $f(y) = \frac{2}{y^2}$, for the interval [1, 2]
- Is this a continuous pdf?
- If so, compute p(1.5 < y < 2.0)
- Graph the pdf using R

Uniform distribution

- The U(0,1) distribution has the function form $p(y) = 1.0$, for $0 < y < 1$; $p(y) = 0$, otherwise.
- Express the following as an integral and compute its value:
  - the probability that a U(0,1) RV is between 0.2 and 0.5
- Using simulation with R, estimate the probability that a U(0,1) RV is between 0.2 and 0.5
- How do the simulated results compare to the calculated results?

Evaluating against the normal distribution

- Summarize the data and examine its statistics
  - summary(), skewness(), kurtosis()
- Histogram
  - hist()
  - a histogram created with hist() has modifiable attributes
    - breaks
    - counts
    - density
  - histograms are often shown with a normal distribution superimposed
- QQ plots
  - qqnorm()
  - qqline()