Statistics

Chapter 4: Probability

Where We’ve Been

- Graphical and Numerical Descriptive Measures for Qualitative and Quantitative Data

Where We're Going

- Probability as a Measure of Uncertainty
- Basic Rules for Finding Probabilities
- Develop the notion of a random variables
- Discrete and continuous random variables and their probability distributions

4.1: Events, Sample Spaces and Probability

- An **experiment** is an act or process of observation that leads to a single outcome that cannot be predicted with certainty.
- A **sample point** is the most basic outcome of an experiment.

4.1: Events, Sample Spaces and Probability

- A **sample space** of an experiment is the collection of all sample points.
  - Roll a single die:

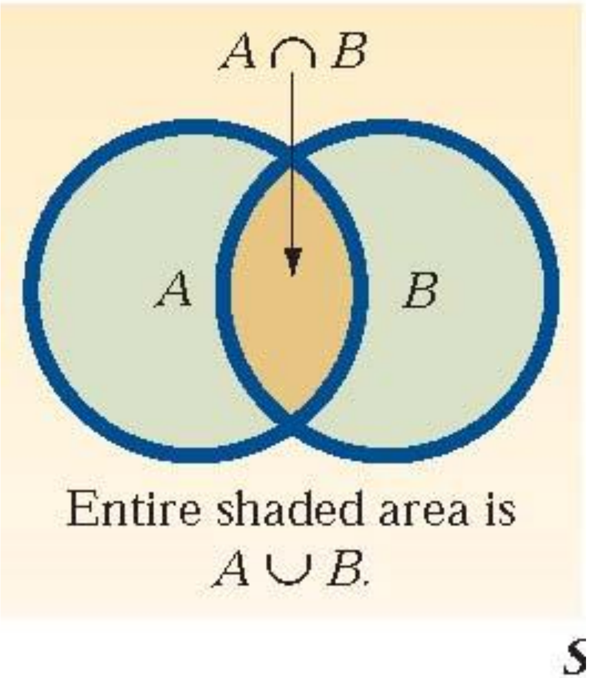
    S: {1, 2, 3, 4, 5, 6}

- An **event** is a specific collection of sample points:
  - Event *A*: Observe an even number.

4.1: Events, Sample Spaces and Probability

- Calculating Probabilities for Events
  - Define the experiment.
  - List the sample points.
  - Assign probabilities to sample points.
  - Collect all sample points in the event of interest.
  - The sum of the sample point probabilities is the event probability.

4.2: Unions, Intersections, Complements



- The **complement** of any event *A* is the event that *A* does not occur, A^C.

P(Ω)=1, P(Φ)=0

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

4.2: Unions, Intersections, Complements

For mutually exclusive events, the probability of the intersection is zero. So

$$P(A \cup B) = P(A) + P(B)$$

In particular, when $B = A^c$ we have

$$P(A^c) = 1 - P(A)$$

4.3: Conditional Probability and Independent Events

- Additional information or other events occurring may have an impact on the probability of an event.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

4.3: Conditional Probability and Independent Events

- The conditional probability formula can be rearranged into the **Multiplicative Rule of Probability** to find joint probability.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B)P(A|B)$$

$$or$$

$$P(A \cap B) = P(A)P(B|A)$$

4.3: Conditional Probability and Independent Events

- Two events A and B are independent if any of the following equivalent conditions is true.

$$P(A \cap B) = P(A)P(B)$$

$$P(A|B) = P(A|B^c) = P(A)$$

$$P(B|A) = P(B|A^c) = P(B)$$

4.4: Random Sampling

- If *n* elements are selected from a population in such a way that every set of *n* elements in the population has an equal probability of being selected, the *n* elements are said to be a (simple) **random sample**.

4.4: Random Sampling

- How many five-card poker hands can be dealt from a standard 52-card deck?
- Use the combination rule:

$$\binom{N}{n} = \binom{52}{5} = \left(\frac{52!}{5!(52-5)!}\right) = 2{,}598{,}960$$

4.4: Random Sampling

- Random samples can be generated by
  - Mixing up the elements and drawing by hand, say, out of a hat (for small populations)
  - Random number generators
  - Random number tables
  - Random sample/number commands on software

4.5: Two Types of Random Variables

- A **random variable** is a variable that assumes numerical values associated with the random outcome of an experiment, where one (and only one) numerical value is assigned to each sample point.
- Function from Ω to R

4.5: Two Types of Random Variables

- A **discrete random variable** can assume a countable number of values.
  - Number of steps to the top of the Eiffel Tower*
- A **continuous random variable** can assume any value along a given interval of a number line.
  - The time a tourist stays at the top once s/he gets there

*Believe it or not, the answer ranges from 1,652 to 1,789. See Great Buildings

4.6: Probability Distributions for Discrete Random Variables

- The **probability distribution** of a discrete random variable is a graph, table or formula that specifies the probability associated with each possible outcome the random variable can assume.
  - $p(x) \geq 0$ for all values of x
  - $\Sigma p(x) = 1$

4.6: Probability Distributions for Discrete Random Variables

- Say a random variable *x* follows this pattern: $p(x) = (.3)(.7)^{x-1}$
  for $x > 0$.
  - This table gives the probabilities (rounded to two digits) for *x* between 1 and 10.

x	*P(x)*
1	.30
2	.21
3	.15
4	.11
5	.07
6	.05
7	.04
8	.02
9	.02
10	.01

4.6: Probability Distributions for Discrete Random Variables

- The **mean**, or **expected value**, of a **discrete random variable** is

$$\mu = E(x) = \sum xp(x).$$

4.6: Probability Distributions for Discrete Random Variables

- The **variance** of a **discrete random variable** *x* is

$$\sigma^2 = E[(x-\mu)^2] = \sum (x-\mu)^2 p(x).$$

- $\sigma^2 = E(X^2) - \mu^2$
- The **standard deviation** of a **discrete random variable** *x* is

$$\sqrt{\sigma^2} = \sqrt{E[(x-\mu)^2]} = \sqrt{\sum (x-\mu)^2 p(x)}.$$

4.6: Probability Distributions for Discrete Random Variables

- In a roulette wheel in a U.S. casino, a \$1 bet on "even" wins \$1 if the ball falls on an even number (same for "odd," or "red," or "black").
- The chance of winning this bet are 47.37%
- 38 slots numbered: 0, 00, 1-36.
- 0, 00 are green. Among the rest, half are red and half are black.
- Place bet on even: 36/2=18 winning possibilities, 20 losing possibilities.
- 18/38=0.4737

4.6: Probability Distributions for Discrete Random Variables

- In a roulette wheel in a U.S. casino, a $1 bet on "even" wins $1 if the ball falls on an even number (same for "odd," or "red," or "black").
- The odds of winning this bet are 47.37%

$$P(win\,\$1) = .4737$$

$$P(lose\,\$1) = .5263$$

$$\mu = +\$1 \cdot .4737 - \$1 \cdot .5263 = -.0526$$

$$\sigma = .9986$$

On average, bettors lose about a nickel for each dollar they put down on a bet like this.

(These are the *best* bets for patrons.)

4.6: Probability Distributions for Discrete Random Variables

Binomial

- The experiment consists of ***n* identical trials.**
- Each trial results in **one of two outcomes**, success (S) or failure (F).
- The probability of success on a single trial is p and **remains constant** from trial to trial. The probability of failure is $q = 1 – p$.
- The trials are **independent**.
- We are interested in ***x*, the number of successes in *n* trials.**

4.6: Probability Distributions for Discrete Random Variables

Binomial

$$P(x=k)=C_k^n p^k q^{n-k}=\frac{n!}{k!(n-k)!}p^k q^{n-k} \text{ for } k=0,1,2,\ldots n.$$

$$C_k^n=\frac{n!}{k!(n-k)!}$$

$$\text{with } n!=n(n-1)(n-2)\ldots(2)1 \text{ and } 0!\equiv 1.$$

$$\text{Mean}: \mu=np$$

$$\text{Variance}: \sigma^2=npq$$

$$\text{Standard deviation}: \sigma=\sqrt{npq}$$

4.6: Probability Distributions for Discrete Random Variables

Geometric

- Each trial results in **one of two outcomes**, success (S) or failure (F).
- The probability of success on a single trial is p and **remains constant** from trial to trial. The probability of failure is $q = 1 – p$.
- The trials are **independent**.
- We are interested in ***x*, the number of trials required to get the first success.**

4.6: Probability Distributions for Discrete Random Variables

Geometric

$$P(X = k) = pq^{\{k-1\}}$$

Mean = $1/p$
Variance = q/p^2

4.6: Probability Distributions for Discrete Random Variables

Poisson

The number of events that occur in a period of time or space during which an average of **μ** such events can be expected to occur.

$$P(x=k)=\frac{\mu^k e^{-\mu}}{k!}$$

For values of k = 0, 1, 2, … The mean and standard deviation of the Poisson random variable are

Mean: μ

Standard deviation: μ

4.6: Probability Distributions for Discrete Random Variables

Hypergemetric

- The experiment consists of ***n* identical trials.**
- Each trial results in **one of two outcomes**, success (S) or failure (F).
- The trials are random draws without replacement from a population of size *N* of which *M* are marked.
- Success is defined as drawing a marked element.
- The random variable **X** is the number of successes.

4.6: Probability Distributions for Discrete Random Variables

Hypergeometric

The probability of exactly *k* successes in *n* trials is

$$P(x = k) = \frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$$

$$\text{Mean}: \mu = n\left(\frac{M}{N}\right)$$

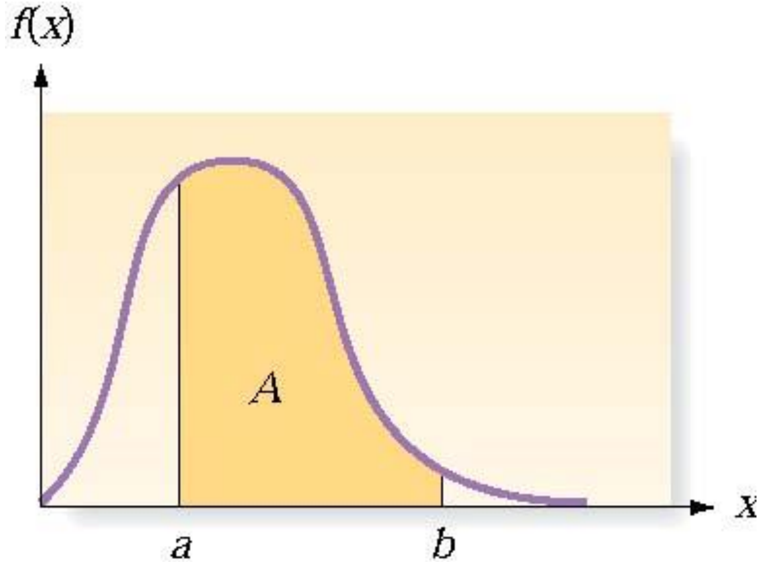
$$\text{Variance}: \sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

4.7: Continuous Probability Distributions

- A **continuous random variable** can assume any numerical value within some interval or intervals.
- The graph of the probability distribution is a smooth curve called a
  - *probability density function*,
  - *frequency function* or
  - *probability distribution.*

4.7: Continuous Probability Distributions

- There are an infinite number of possible outcomes
  - p(*x*) = 0
  - Instead, find p(a<*x*<b)
    - ☺ Table
    - ☺ Software
    - ☹ Integral calculus)



4.7: Continuous Probability Distributions

- Uniform (special case of beta)
- Exponential (special case of gamma)
- Gamma (support on half real line)
- Beta (bounded support)
- Normal (support on R)

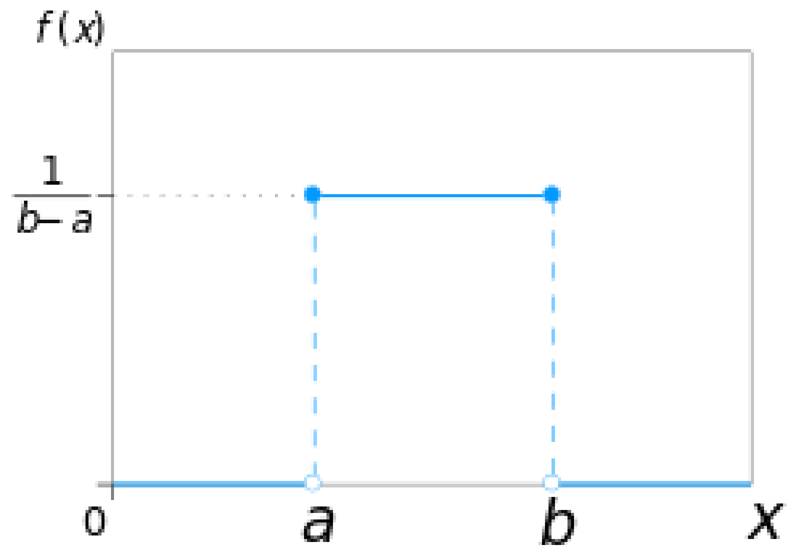
4.7: Continuous Probability Distributions

- Uniform

Pdf: $f(x)=1/(b-a),\ a<x<b$

Mean $(a+b)/2$

Var $(b-a)^2/12$



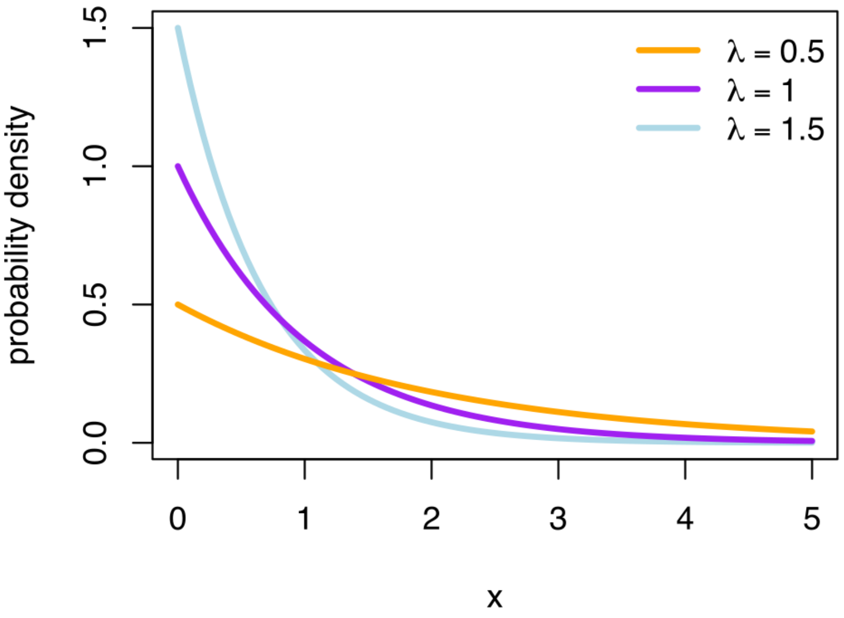
4.7: Continuous Probability Distributions

- Exponential

Pdf: $f(x)=\lambda \exp(-\lambda x)$, $x>0$

Mean $1/\lambda$

Var $1/\lambda^2$



4.7: Continuous Probability Distributions

- Gamma

Pdf : $\dfrac{\beta^{\alpha} x^{\{\alpha-1\}} \exp(-\beta x)}{\Gamma(\alpha)}, x > 0$

Mean $\dfrac{\alpha}{\beta}$

4.7: Continuous Probability Distributions

- Beta

Pdf $\frac{x^{\{\alpha-1\}}(1-x)^{\{\beta-1\}}}{\mathrm{B}(\alpha,\beta)}, 0 < \mathrm{x} < 1$

Mean $\frac{\alpha}{\alpha+\beta}$

4.8: The Normal Distribution

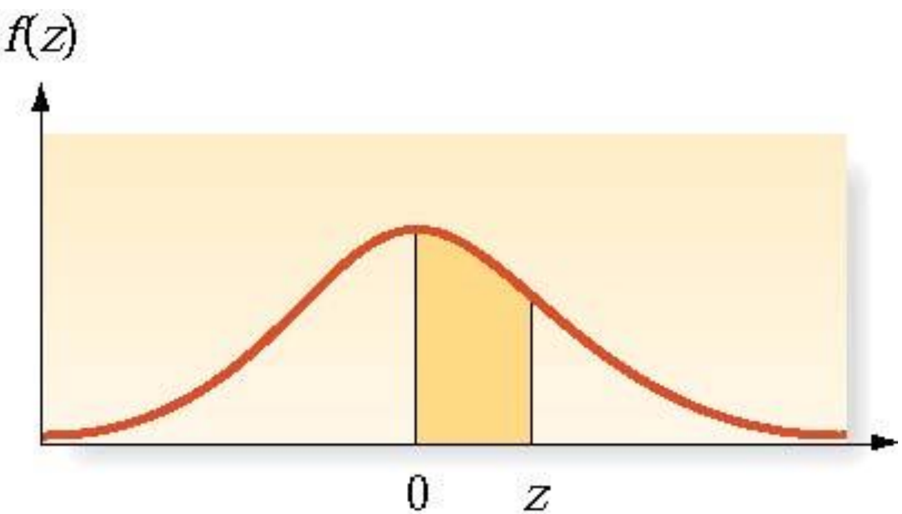
- Closely approximates many situations
  - Perfectly symmetrical around its mean
- The probability density function *f(x):*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[(x-\mu)/\sigma]^2}{2}}$$

μ = the mean of x

$\boldsymbol{\sigma}$ = the standard deviation of x

4.8: The Normal Distribution



$$P(0 < z < 1.00) \quad = .3413$$

$$P(-1.00 < z < 0) = .3413$$

$$P(-1 < z < 1) \quad = .3413 + .3413$$
$$= .6826$$

$$P(1 < z < 1.25) \quad =$$
$$P(0 < z < 1.25) - P(0 < z < 1.00)$$
$$= .3944 - .3413 = .0531$$

4.8: The Normal Distribution

For a normally distributed random variable *x*, if we know μ and **σ**, z follows normal with mean 0 and variance 1

$$z_i = \frac{x_i - \mu}{\sigma}$$

So *any* normally distributed variable can be analyzed with this single distribution