ADVANCED BIOSTATISTICS ABSTAT17

Quick review of Statistical Concepts

Carina Silva
(*carina.silva@estesl.ipl.pt*)
Escola Superior de Tecnologias da SaÃºde de Lisboa &
Center of Statistics and Applications, University of Lisbon

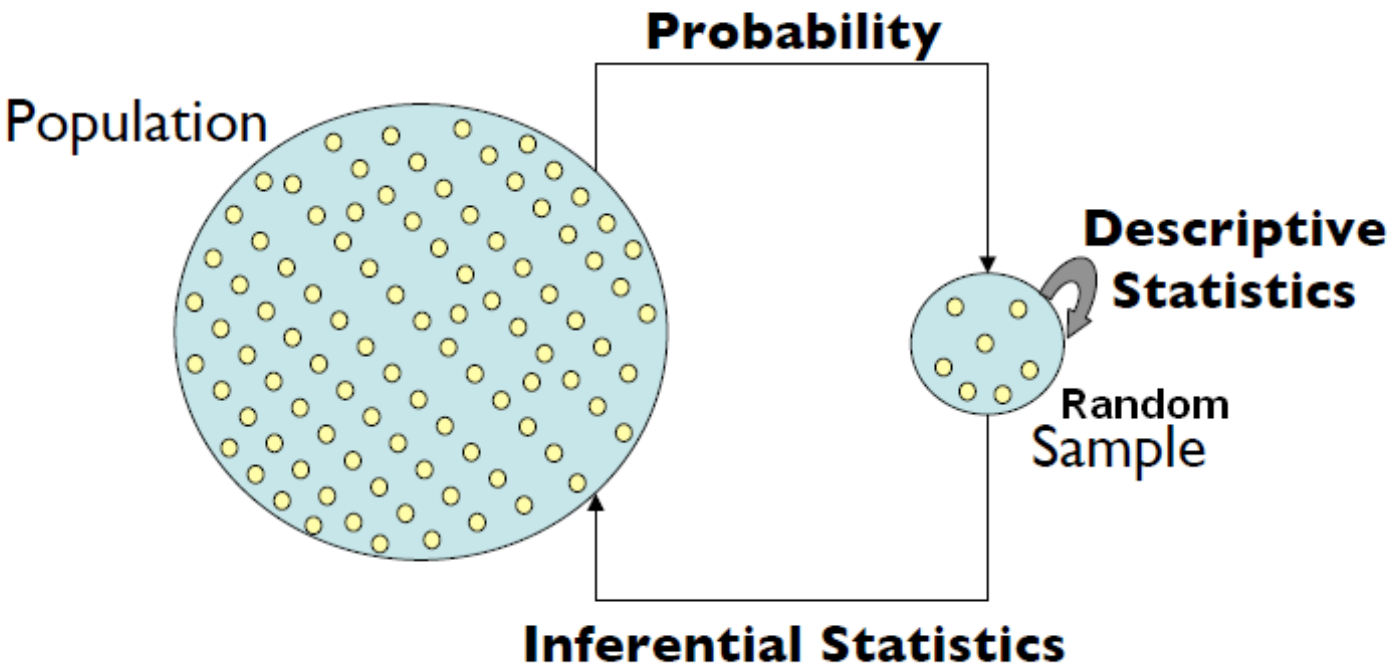
IGC, April, 10–13, 2017

Introduction

Probability used to describe situations were uncertainty occurs.

Our main objective: develop the art of describing uncertainty in terms of probabilistic models.

The "central dogma"of inferential statistics



Introduction

Deterministic models: The equations can be used to determine the value of a specific variable in the model based on the knowledge of the values assumed by other model variables.

But there is always some uncertainty in experimental science.

Thus we need:
Statistical models: Allow us to assess the degree of uncertainty present in our experimental results.

Probability

The mathematics on which statistical methods rest is probability theory.

We will have to start introducing some basic ideas of probability.

Methodology	Probability viewpoint	Source of Information
Frequentist (Classical)	Frequentist	Sample data
Bayesian	Subjective	Prior and sample data

Probability - Kolmogorov's Axioms of Probability

$\Omega =$ sample space, set of all possible outcomes.
$\mathcal{A} =$ set of events
$P =$ The assignment of probabilities to the events

- **Definition of probability**:
  Is a set function $P : \mathcal{A} \to [0, 1]$ which satisfies the axioms
  1. $P[A] \geq 0$ for every $A \in \mathcal{A}$.
  2. $P[\Omega] = 1$.
  3. $A_1, A_2, ...$ sequence of mutually exclusive events in $\mathcal{A}$, then

$$P\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} P[A_i] = P[A_1] + P[A_2] + ...$$

- **Definition of Probability space**
  Is the triplet $(\Omega, \mathcal{A}, P[.])$

Conditional Probability

Let $(\Omega, \mathcal{A}, P[.])$ be a probability space

Definition: Conditional Probability

Let A and B two events in $\mathcal{A}$, $P[B] > 0$. Conditional probability of A given B

$$P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

Independence

Definition: Independent events

A and B two events in $\mathcal{A}$ are independent if one of the following is satisfied

(i) $P[A \cap B] = P[A]P[B]$

(ii) $P[A|B] = P[A]$ if $P[B] > 0$

(iii) $P[B|A] = P[B]$ if $P[A] > 0$

Remark: Independent events can only be mutually exclusive if the probability of at least one is zero.

Let $B_1, ..., B_n$ mutually disjoint (or mutually exclusive) in $\mathcal{A}$, $\Omega = \bigcup_{i=1}^{n} B_i$ and $P[B_i] > 0, \forall i$

Theorem - Theorem of total probabilities

For every $A \in \mathcal{A}$

$$P[A] = \sum_{i=1}^{n} P[A|B_i]P[B_i].$$

Theorem - Bayes' Theorem

For every $A \in \mathcal{A}$ such that $P[A] > 0$

$$P[B_k|A] = \frac{P[A|B_k]P[B_k]}{\sum_{i=1}^{n} P[A|B_i]P[B_i]}.$$

Both theorems remain true if $n = \infty$.

Basic Concepts

- **Variable**: is a characteristic or condition that changes or has different values for different individuals.

Basic Concepts

- **Variable**: is a characteristic or condition that changes or has different values for different individuals.
- **Observation** (or case): Is a realization of a variable. For example, the weight of a randomly chosen rat is such an observation. (Observations are represented by lowercase, e.g. x_1, x_2.)

Basic Concepts

The following definitions are vital in understanding descriptive statistics:

Types of Data

Qualitative or Categorical

- Nominal (race, blood type, etc.)
- Ordinal (the order or rank of the categories is meaningful. For example, staff members may be asked to indicate their satisfaction with a training course on an ordinal scale ranging from "poor" to "excellent".)

Quantitative

- Discrete (number of planets orbiting a distant star. Could even be countably infinite.)
- Continuous (your age, not rounded off; weigth)

Random Variable

A **Random Variable** is a function, which assigns unique numerical values to all possible outcomes of a random experiment under fixed conditions.

Example:

Suppose that a coin is tossed three times and the sequence of heads and tails is noted. The sample space for this experiment evaluates to:
$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Now let the random variable X be the number of heads in three coin tosses. X assigns each outcome in S a number from the set $Sx = \{0, 1, 2, 3\}$. The table below lists the eight outcomes of S and the corresponding values of X.

Outcome	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
X	3	2	2	2	1	1	1	0

Random Variable

The condition for a function to be a random variable is that the random variable cannot be multivalued.

There are two types of random variables:

Discrete Random Variable is one that takes a finite distinct values (countable number of possibilities). Example: X- A number of students who fail a test, $P(X = k)$.

Continuous Random Variable is one that takes an infinite number of possible values (can take any value on the real line or some subset of the real line). Example: X - Duration of a call in a telephone exchange, $P(X \leq 10)$.

We distinguish discrete from continuous random variables according to whether the sample space S is countable or not countable.

Basic Concepts

Parameter: is a numeric quantity, usually unknown, that describes a certain population characteristic. For example, the population mean, μ, is a parameter that is often used to indicate the average value of a quantity. (Usually are represented by Greek letters, e.g. μ, σ.)

Basic Concepts

- **Estimator/Statistic**: is a statistic (that is, a function of the data) that is used to infer the value of an unknown parameter in a statistical model. Suppose there is a fixed parameter θ that needs to be estimated. An estimator of θ is usually denoted by the symbol $\hat{\theta}$. If X is used to denote a random variable corresponding to the observed data, the estimator (itself treated as a random variable) is symbolized as a function of that random variable, $\hat{\theta}(X)$.

Basic Concepts

- **Estimate/Statistic**: An estimate is the particular value of an estimator that is obtained from a particular sample of data and used to indicate the value of a parameter. (e.g. $\bar{x}$ is an estimate of μ). The estimate for a particular observed data set (i.e. for $X = x$) is then $\hat{\theta}(x)$, which is a fixed value.

Probability Mass Function (pmf)

For a discrete random variable (r.v.) the probability distribution of X is completely determined by specifying $P_X[E]$ for $E = \{x\}$, for every $x \in S$.
We represent $P[X = x]$ by $f(x)$. Remember that S is countable.
It is a function from $\Re$ to $[0, 1]$ such that

1. $f(x) = \begin{cases} \geq 0, & x \in S; \\ 0, & \text{otherwise.} \end{cases}$
2. $\sum_{x \in S} f(x) = 1$

To distinguish the different values x in S we can write x_i.

Moments of Discrete Random Variables

Definition:*Mean of a discrete r.v.*

The mean (expected value) of a discrete random variable X with sample space S and p.m.f. $p(x)$ is

$$E[X] = \sum_{x \in S} xp(x).$$

- Usually the mean is represented by μ.
- When we do not know the numerical values of $p(x)$ (which is the common situation) the mean is not known. We say that the mean is a parameter.
- The mean is not necessarily a realizable value of a discrete r.v.
- The expressions "mean of the r.v. X"and "mean of the probability distribution of the r.v. X"are equivalent.

Moments of Discrete Random Variables

- The concept of the mean of a r.v. X can be generalized to the mean of any function (a r.v. as well) $g(X)$ of X, as

$$E[g(X)] = \sum_{x \in S} g(x)p(x).$$

- Linearity property: If X is a r.v. and a and b are constants, then $E[aX + b] = aE[X] + b$.
- If $X_1, ..., X_n$ is a sequence of r.v. then

$$E[X_1 + ... + X_n] = E[X_1] + ... + E[X_n].$$

- However $E[X_1...X_n] = E[X_1]...E[X_n]$ if the r.v. are independent.

Definition: *Variance of a discrete r.v.*

The variance of a discrete r.v. X with mean μ is

$$var(X) = E[(X - \mu)^2] = \sum_{x \in S} (x - \mu)^2 f(x).$$

- The usual notation for the variance is σ^2.
- The variance is always non-negative.
- The variance is a measure of dispersion of the probability distribution of the r.v. around its mean.
- The expressions "variance of the r.v. X"and "variance of the probability distribution of the r.v. X"are equivalent.

Moments of Discrete Random Variables

- The standard deviation is the positive root of the variance. We use the notation σ for it.
- If σ^2 is the variance of a random variable X and a and b are constants, then $var(a + bX) = b^2\sigma^2$.
- If μ and σ^2 are, respectively, the mean value and the variance of a r.v. X, then

$$\sigma^2 = E[X^2] - \mu^2.$$

Cumulative distribution function

The **cumulative distribution function** (cdf), or just **distribution function**, describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to x. Intuitively, it is the "area so far"function of the probability distribution.

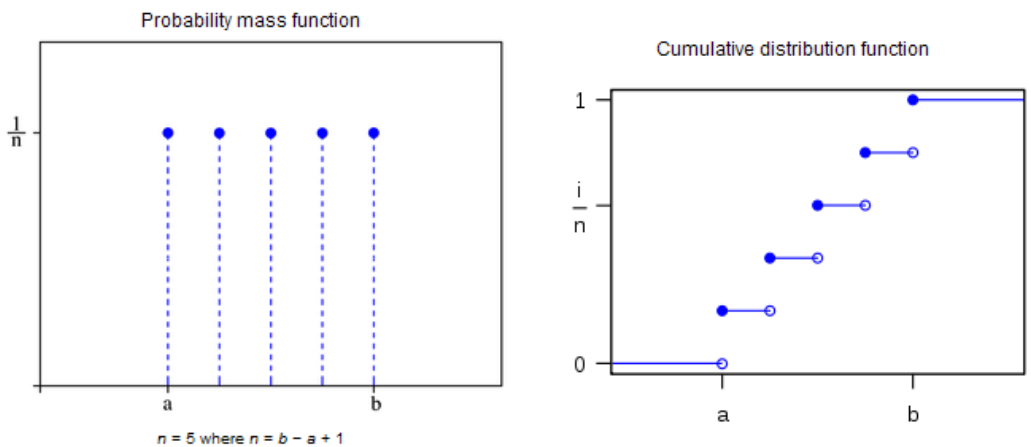
Discrete cdf

$$P(X \leq x) = \sum_{i=1}^{x} f(x)$$

Discrete Uniform Distribution

The **uniform distribution** $U\{a, b\}$
has the same value at each point on the domain. This distribution is often used to express an unwillingness to make a choice or a lack of information.

$P(X = x) = \frac{1}{k+1}$, for all values of $x = 0, 1, 2, ..., k$



- **Bernoulli distribution** $Ber(1, \theta)$,
  parameter ($\theta \in [0, 1]$).
  We call "Bernoulli trial"a single trial with two possible outcomes "success"and "failure".
  The distribution of the random variable associated to this trial is called Bernoulli distribution. θ is the probability of success.
  support $S = \{0, 1\}$

$$P[X = x|\theta] = p(x|\theta) = \begin{cases} \theta^x(1-\theta)^{1-x}, & \text{for } x = 0, 1; \\ 0, & \text{otherwise.} \end{cases}$$

Verify that

$$E[X] = \theta \quad var[X] = \theta(1-\theta).$$

Bernoulli distribution

Example: For example, if we are studying a certain genetic disease and are observing one individual to determine if they have it, we might consider it a "success"if they do have it.

Let's say it is known that 2.6% of the population has the disease. Then when we randomly select one person and observe their disease status. We define X to be 1 if they have it and 0 if they don't. Then the pmf of X is:

$$P(X = x) = \begin{cases} 1 - 0.026, & \text{for } x = 0; \\ 0.026, & \text{for } x = 1. \end{cases}$$

Binomial Distribution

The binomial distribution describes the behavior of a count variable X if the following conditions apply:

1: The number of observations n is fixed.
2: Each observation is independent.
3: Each observation represents one of two outcomes ("success"or "failure").
4: The probability of "success"p is the same for each outcome.

Binomial Distribution

Binomial distribution $Bi(n, \theta)$,

parameter ($\theta \in [0, 1]$)

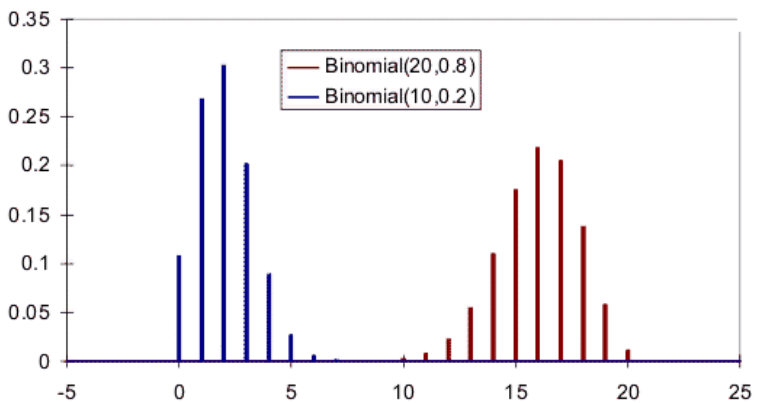
The random variable which counts the number of successes in n independent and identical Bernoulli trials is called Binomial random variable.

support $S = \{0, 1, ..., n\}$

$$P[X = x|\theta] = p(x|\theta) = \begin{cases} \binom{n}{x}\theta^x(1-\theta)^{n-x}, & \text{for } x \in S; \\ 0, & \text{otherwise.} \end{cases}$$

Verify that

$$E[X] = n\theta, \quad var[X] = n\theta(1-\theta).$$



Binomial distribution

Example:In our disease example, let's say again that we know the proportion in the population is 2.6%. Then let's randomly select 10 people who are independent of one another. For each of the 10 we will observe whether or not they have the disease. X is the r.v. who represents the number of the 10 that have the disease, which is a Binomial random variable. In this example, the θ in the formula is .026 and the n is 10. So for this specific X, its pmf is:

$$P(X = x) = \binom{10}{x}(.026)^x(1 - 0.026)^{10-x}, x = 0, ..., 10$$

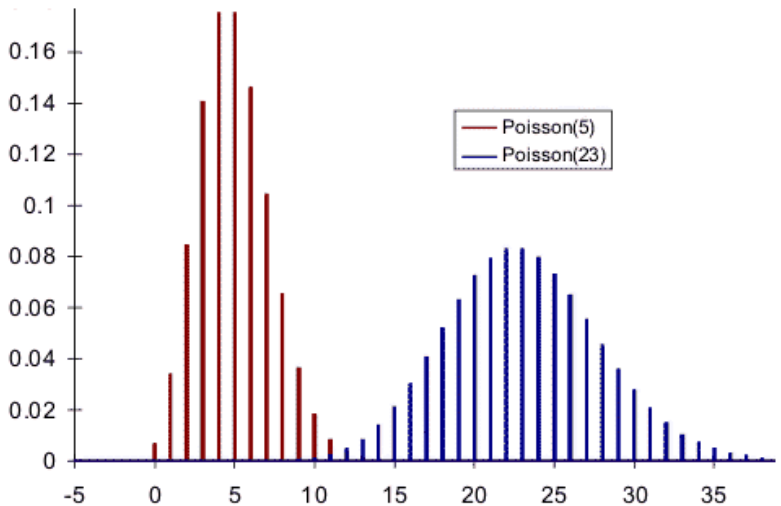
The chance that we will find that two of the ten have the disease is $P(X = 2) = \binom{10}{2}(.026)^2(1 - .026)^8 = .0246.$

Poisson Distribution

The **Poisson distribution** is usually used to model the number of events occurring within a given time interval.

$$X \cap P(\lambda)$$

$$P(X = x) = \frac{\lambda^x e^{-x}}{x!}, \text{ for } x = 0, 1, 2, 3...$$



Probability Density Function - pdf

- A probability density function (pdf), is a function associated with a **continuous random variable**
- Areas under pdfs correspond to probabilities for that random variable
- To be a valid pdf, a function f must satisfy
  1. $f(x) \geq 0$ for all x
  2. $\int_{-\infty}^{+\infty} f(x)dx = 1$

Cumulative Distribution Function - cdf

- **Definition**: *Cumulative distribution function (cdf)*
  $F_X : \Re \to [0, 1]$, satisfying

$$F_X(x) = P[X \leq x]$$

  for every real x.

F_X describes the probability that a random variable X with a given probability distribution will be found at a value less than or equal to x.

Uniform Distribution

One of the most important applications of the uniform distribution is in the generation of random numbers. That is, almost all random number generators generate random numbers on the (0,1) interval.

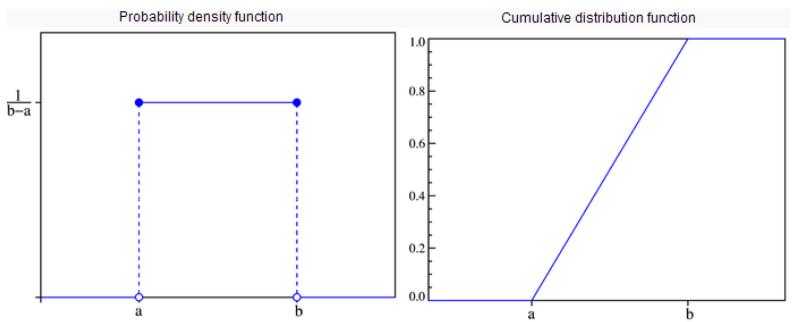
- Uniform $U(a, b)$, $\quad a < b$

$$f_X(x) = \frac{1}{b-a}, \quad x \in [a, b]$$

$$E[X] = \frac{b+a}{2}, \quad var[X] = \frac{(b-a)^2}{12}$$

Uniform Distribution

The Uniform distribution assigns equal probabilities to all possible ranges of equal length within which the r.v. can fall. This distributions is widely used in bioinformatics, Bayesian analysis, quantitative genetics and so on.



Normal distribution

- Normal (Gaussian) $N(\mu, \sigma^2)$. $(\mu \in \Re, \sigma^2 > 0)$

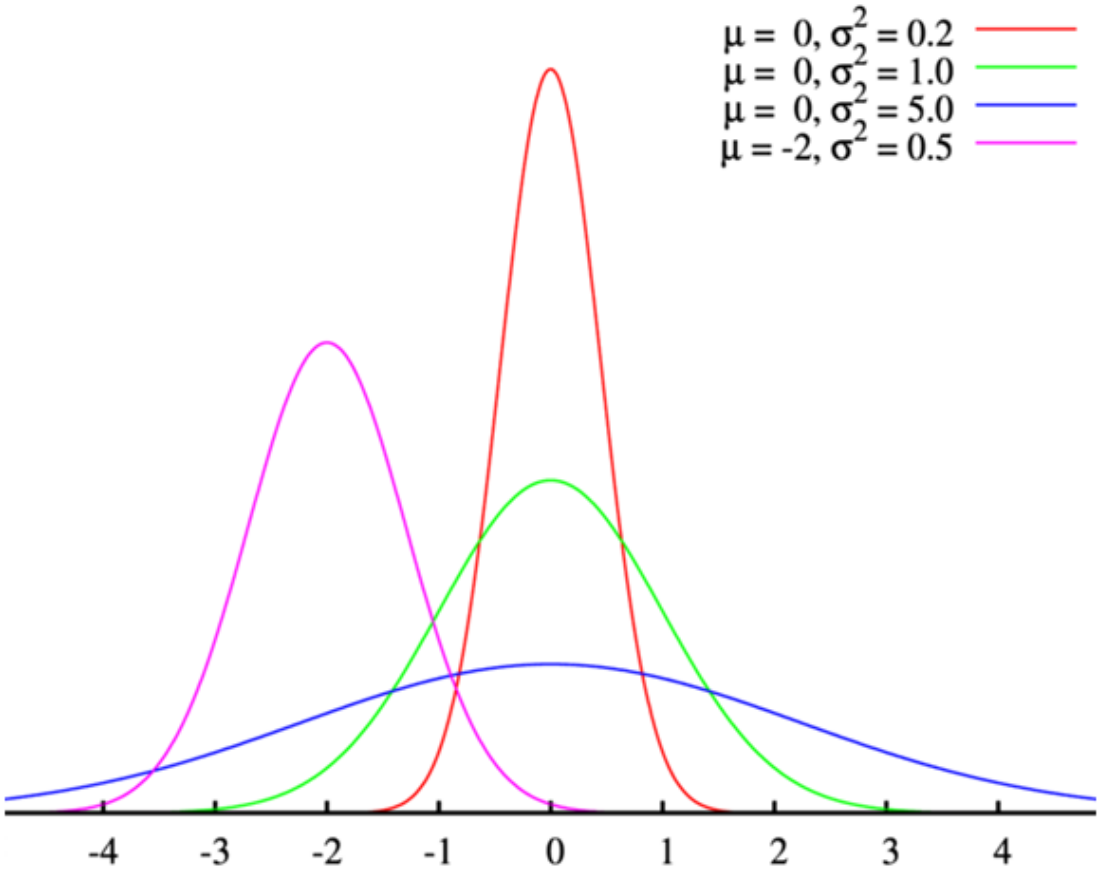
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < +\infty.$$

$$E[X] = \mu, \quad var[X] = \sigma^2.$$

μ is a location parameter and σ^2 is a scale parameter.

Note constants:
$\pi = 3.14159$
$e = 2.71828$

Normal distribution

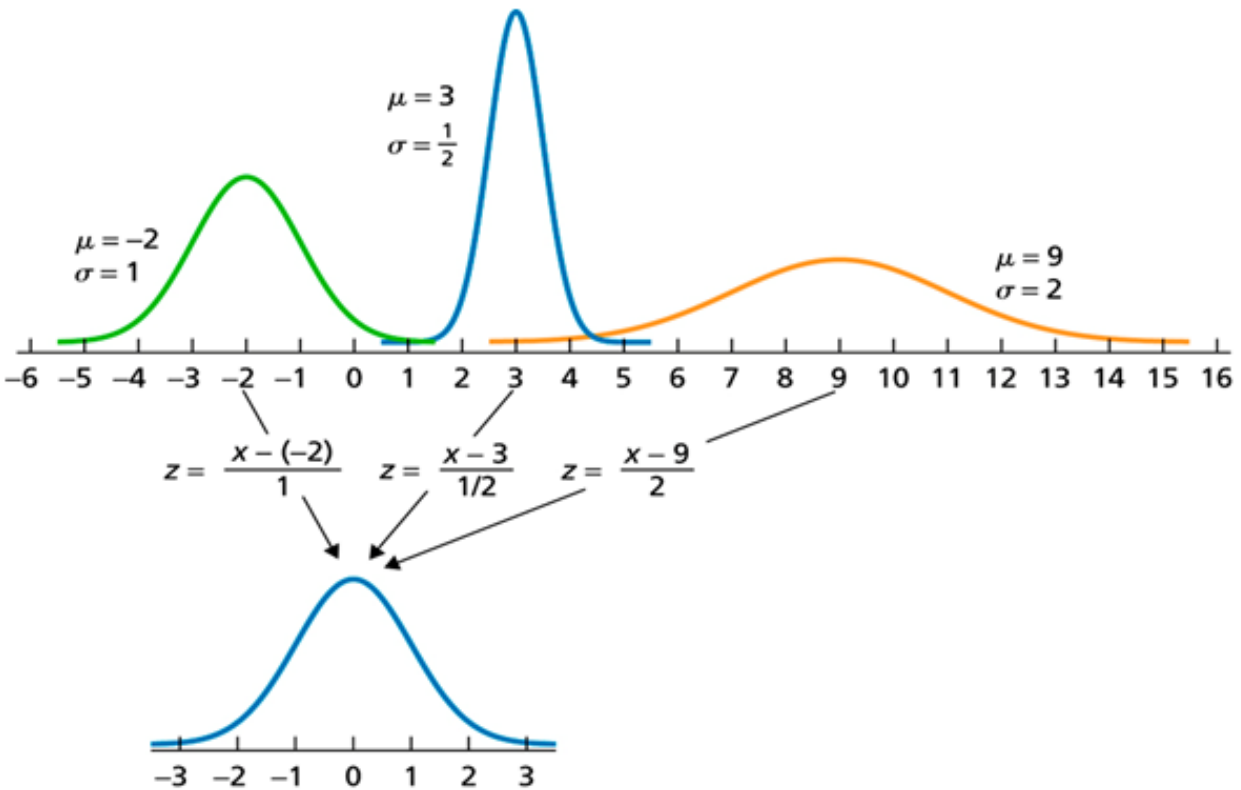


Standard Normal Distribution

All normal distributions can be converted into the standard normal curve by subtracting the mean and dividing by the standard deviation:

$$Z = \frac{X - \mu}{\sigma}$$

then Z has a Normal(0, 1) distribution. Such Z is called a standard normal random variable. The scale of Z has no units and it is called the standardized scale. It shows how many standard deviations from the mean the X variable is.


$\mu = 3$
$\sigma = \frac{1}{2}$
$\mu = -2$
$\sigma = 1$
$\mu = 9$
$\sigma = 2$
−6 −5 −4 −3 −2 −1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
$z = \frac{x-(-2)}{1}$
$z = \frac{x-3}{1/2}$
$z = \frac{x-9}{2}$
−3 −2 −1 0 1 2 3

Multiple choice questions

Suppose that the probability of event A is 0.2 and the probability of event B is 0.4. Also, suppose that the two events are independent. Then P(A|B) is:

A. $P(A) = 0.2$

B. $P(A)/P(B) = 0.2/0.4 = 1/2$

C. $P(A) \times P(B) = (0.2)(0.4) = 0.08$

D. None of the above.

If two events (both with probability greater than 0) are mutually exclusive, then:

A. They also must be independent.

B. They also could be independent.

C. They cannot be independent.

Just one option is correct.

A mass probability function is a rule of correspondence or equation that:

a) Finds the mean value of the random variable.
b) Assigns values of x to the events of a probability experiment.
c) Assigns probabilities to the various values of x.
d) Defines the variability in the experiment.
e) None of the above is correct.

Q: A Z-Score is a

a. raw score with a mean of zero;
b. raw score with a mean of 50;
c. standard score with a mean of zero;
d. standard score with a mean of 50.

Q: Z-scores provide information about the location of raw scores
a. below the mean in units of the range of the distribution;
b. above the mean in units of the standard deviation of the distribution;
c. above and below the mean in units of the range of the distribution;
d. above and below the mean in standard deviation units from the mean.

Probability Distributions in R

Different distributions can be easily calculated or simulated using R. The functions are named such that the first letter states what it calculates or simulates:
d=density function, p=distribution function, q=quantile,
r=random generation,
and the last part the function's name specifies the type of distribution:
unif=uniform, binom=binomial, pois=poisson

Probability Distributions with R

Calculating the probability density function: To calculate the value of the p.d.f. for a N(2,25) using the quantile, x, type:

```
> dnorm(x, mean=2, sd=5)
```

Calculating the cumulative density function: To calculate the value of the c.d.f. for a N(2, 25) using the quantile, x, type:

```
> pnorm(x, mean=2, sd=5)
```

Determining a quantile: To calculate the quantile associated with a N(2, 25) using the probability, x, type:

```
> qnorm(x, mean=2, sd=5)
```

Generating a random value from a distribution: To generate 10 random values from a N(2, 25), type:

```
> rnorm(10, mean=2, sd=5)
```

Probability Distributions with R

Different distributions can be easily calculated or simulated using R. The functions are named such that the first letter states what it calculates or simulates

d = density function, p = distribution function, q = quantile,
r = random generation,

and the last part of the function's name specifies the type of distribution

beta = beta,	chisq = chi-squared,	exp = exponential,
f = fisher,	gamma = gamma,	logis = logistic,
norm = normal,	t = student,	unif = uniform,
weibull = weibull,	binom = binomial,	nbinom = negative bin.,
pois = poisson.		

For instance, the function, qnorm, returns the quantiles of the normal distribution.

Quantiles

The α^{th} **quantile** of a distribution with cumulative distribution F is the point x_α:
$F(x_\alpha) = \alpha$
A **percentile** is simply a quantile with α expressed as a percent.
The **median** is the 50^{th} percentile.

Random Vectors

Random vectors are simply random variables collected into a vector

- For example if X and Y are random variables, (X, Y) is a random vector.

Useful fact

We will use the following fact extensively in this class: If a collection of random variables $X_1, X_2, ..., X_n$ are independent, then their joint distribution is the product of their individual densities or mass functions. That is, if f_i is the density for random variable X_i we have that
$f(x_1, x_2, ..., x_n) = \Pi_{i=1}^{n} f_i(x_i)$

IID random variables

- If we have $f_1 = f_2 = ... = f_n$, we say that the random variables X_i are iid for independence and identically distributed.
- iid random variables are the default model for random samples

Some comments

- When X_i are independent with a common variance, σ^2: $Var(\bar{X}) = \frac{\sigma^2}{n}$
- $\frac{\sigma}{\sqrt{n}}$ is called the **standard error** of the sample mean
- The standard error of the sample mean is the standard deviation of the distribution of the sample mean
- σ is the standard deviation of the distribution of a single observation
- Easy way to remember, the sample mean has to be less variable than a single observation.

Avoiding some confusion

- Suppose X_i are iid with mean μ and variance σ^2
- S^2 estimates σ^2
- $S/\sqrt{n}$ estimates $\sigma/\sqrt{n}$ the standard error of the mean
- $S/\sqrt{n}$ is called the sample standard error (of the mean)

Define likelihood

- A common and fruitful approach to statistics is to assume that the data arises from a family of distributions indexed by a parameter that represents a useful summary of the distribution
- The likelihood of a collection of data is the joint density evaluated as a function of the parameters with the data fixed
- Likelihood analysis of data uses the likelihood to perform inference regarding the unknown parameter

Likelihood

Given a statistical probability mass function or density, say $f(x, \theta)$, where θ is an unknown parameter, the likelihood is f viewed as a function of θ for a fixed, observed value of x.

Interpretations of likelihoods

- Ratios of likelihood values measure the relative evidence of one value of the unknown parameter to another.
- Given a statistical model and observed data, all of the relevant information contained in the data regarding the unknown parameter is contained in the likelihood.
- If $\{X_i\}$ are independent events, then their likelihoods multiply. That is, the likelihood of the parameters given all of the X_i is simply the product of the individual likelihoods.

Example

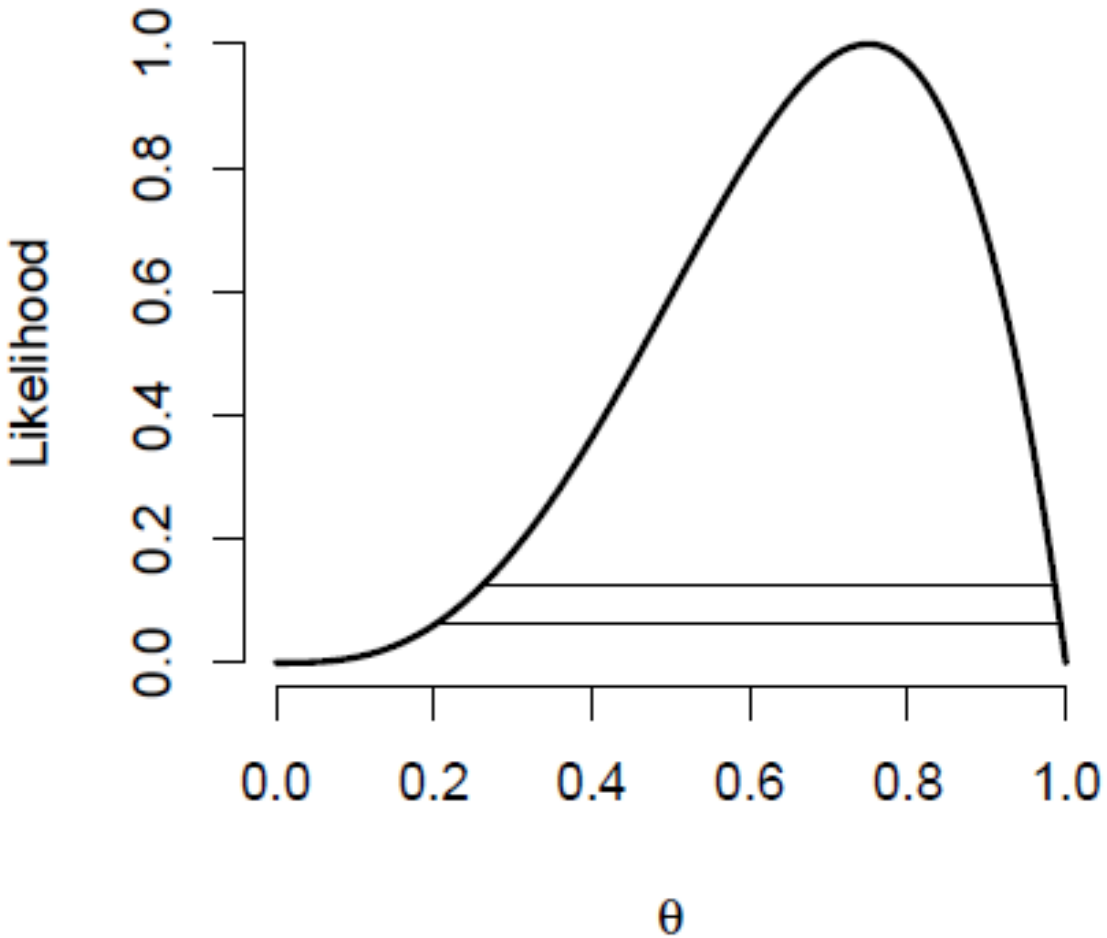
- Suppose that we flip a coin with success probability θ
- Recall that the mass function for x $f(x, \theta) = \theta^x(1-\theta)^{1-x}$ for $\theta \in [0, 1]$
  where x is either 0 (Tails) or 1 (Heads)
- Suppose that the result is a Head
- The likelihood is:
  $\mathcal{L}(\theta, 1) = \theta^1(1-\theta)^{1-1} = \theta$ for $\theta \in [0, 1]$
- Therefore, $\mathcal{L}(.5, 1/\mathcal{L}(.25, 1) = 2$
- There is twice as much evidence supporting the hypothesis that $\theta = 0.5$ to the hypothesis that $\theta = 0.25$

Example (cont)

- Suppose now that we flip our coin from the previous example 4 times and get the sequence 1, 0, 1, 1
- The likelihood is:
  $\mathcal{L}(\theta, 1, 0, 1, 1) =$
  $\theta^1(1-\theta)^{1-1}\theta^0(1-\theta)^{1-0}\theta^1(1-\theta)^{1-1}\theta^1(1-\theta)^{1-1} = \theta^3(1-\theta)^1$
  for $\theta \in [0, 1]$

Plotting likelihoods

- Generally, we want to consider all the values of θ between 0 and 1
- A likelihood plot displays θ by $\mathcal{L}(\theta, x)$
- Usually, it is divided by its maximum value so that its height is 1


Likelihood
0.0
0.2
0.4
0.6
0.8
1.0
θ

Maximum likelihood

- The value of θ where the curve reaches its maximum has a special meaning
- It is the value of θ that is most well supported by the data
- This point is called the maximum likelihood estimate (or MLE) of θ:
  $MLE = argmax_\theta \mathcal{L}(\theta, x)$
- Another interpretation of the MLE is that it is the value of θ that would make the data that we observed most probable.

Acknowledgements:
Antónia Turkman (DEIO-FCUL), for allowing the use of some material produced by us in previous courses.