

Modèles stochastiques II

INFO 154

Gianluca Bontempi

Département d'Informatique
Boulevard de Triomphe - CP 212
<http://www.ulb.ac.be/di>

The basics of statistics

Statistics starts with a problem, continues with the collection of data, proceeds with the data analysis, and finishes with conclusions.

Phenomenon: e.g. the weather in Belgium;

Variables: e.g. the temperature.

Observations: e.g. the historical recording of temperature in the last ten years.

Data: formatted version of observations (management of data entry errors, missing data, re-normalization, scaling, coding).

Data analysis methods: e.g., time series prediction.

Inferences or conclusions: e.g. *the temperature's planet is gradually getting hotter.*

Variability in measurements

- Consider an unknown and discrete scalar variable $z \in \mathcal{Z}$, (e.g. the age of a student inscribed to the last year of Computer Science).
- If we measures N times this variable we realize that this variable is affected by uncertainty.
- We often use simple **graphical methods** to assist in analyzing the data from an experiment.

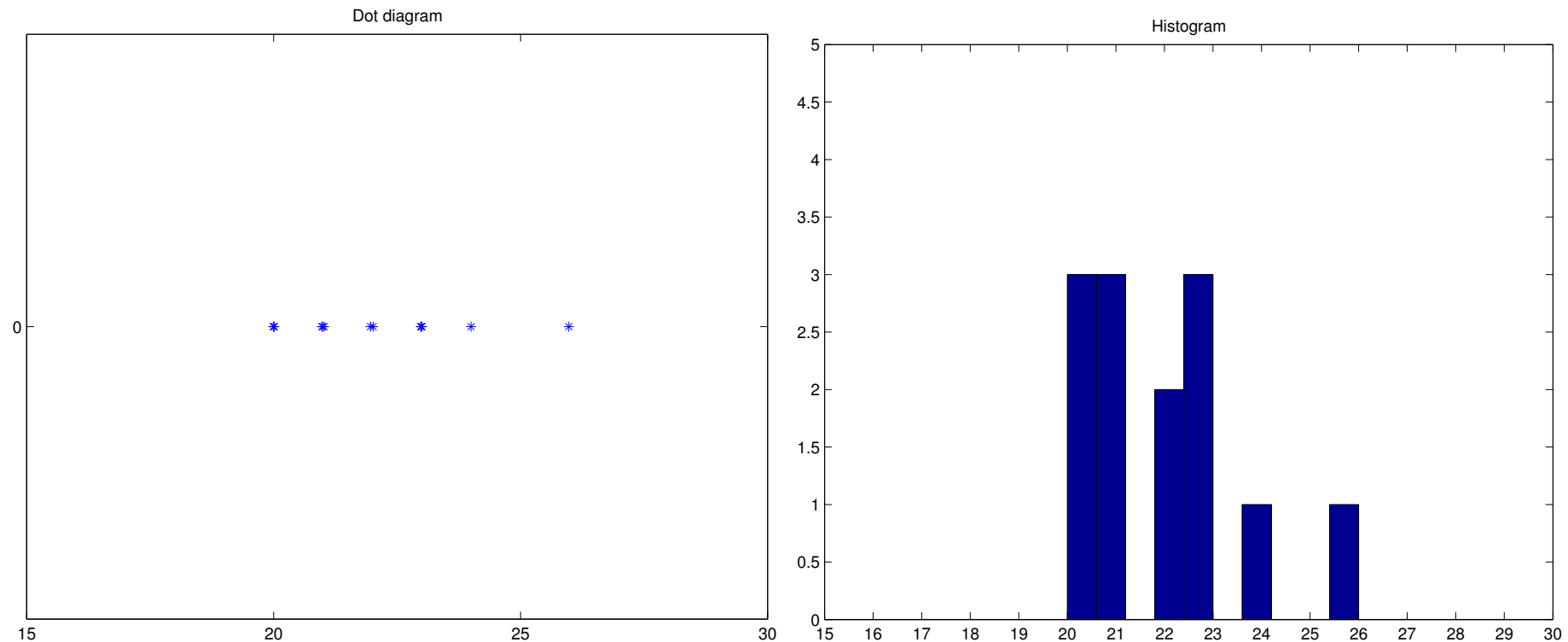
Dot diagram: this enables to see quickly the general location or central tendency of the observations and their spread. useful for displaying a small body of data.

Histogram: it shows the central tendency, spread and general shape of the distribution of a single variable starting from a sample data. It is constructed by dividing the horizontal axis into H intervals (usually of equal length) and drawing a rectangle over the j th interval with area of the rectangle proportional to N_j , the number of observations that fall in that interval

Visualization of variability

Set of measurements

$$D_N = \{20, 23, 24, 21, 26, 22, 23, 20, 21, 23, 22, 21, 20\}$$



The histogram is constructed by dividing the horizontal axis into $H = 10$ intervals.

To see the effect of the number H on the visualization of 100 measures, run the MATLAB script `s_hist.m`.

Table representation

For a large number n of variables, the graphical representation is uneasy. In this case, tables are currently used.

Samples	Variables				
1	z_{11}	...	z_{ij}	...	z_{1n}
⋮	⋮	⋮	⋮	⋮	⋮
i	z_{i1}	...	z_{ij}	...	z_{in}
⋮	⋮	⋮	⋮	⋮	⋮
N	z_{N1}	...	z_{Nj}	...	z_{Nn}

- If all the variables are quantitative, the table is a matrix.
- Each line represents a sample (e.g. an individual) for which we have measures n different features (e.g. student's age, student's height, student's income).
- Each column represents the N different measures collected for a specific variable (the N ages of the students in the classroom).

Probabilistic interpretation of uncertainty

- This course will **assume** that the variability of measurements can be **represented** by the probability formalism.
- In the case of our discrete variable z , we assume that this is a **discrete random variable**.
- A **discrete random variable** is a numerical quantity, linked to some experiment involving some degree of randomness, that takes its value from some discrete set of possible values.
- Example: the experiment might be the rolling of two six-sided dice and the r.v. z might be the sum of the two numbers showing in the dice. In this case the set of possible values are $2, \dots, 12$.
- In the example of the age, the random experiment is a **compact (and approximate)** way of modeling the disparate set of causes which led to variability in the value of z .

Probability distribution of a discrete r.v.

The **probability distribution** of a discrete r.v. z is the combination of

1. the set \mathcal{Z} of values that this r.v. can take (also called **range** or **sample space**)
2. the set of probabilities associated to each value of \mathcal{Z}

This means that we can attach to the random variable some specific mathematical function $P(z)$ that gives for each $z \in \mathcal{Z}$ the probability that z assumes the value z

$$P_{\mathbf{z}}(z) = \text{Prob} \{ \mathbf{z} = z \}$$

Probability distribution of a discrete r.v.(II)

For a reduced number of possible values of z , the probability distribution can be presented in the form of a table. For example, if we plan to toss a coin twice, and the random variable z is the number of heads that eventually turn up, the probability distribution can be presented as follow

Values of the random variable z	0	1	2
Associated probabilities	0.25	0.50	0.25

Parametric distribution function

Suppose that

1. z is a discrete r.v. that takes its value in $\mathcal{Z} = \{1, 2, 3\}$.
2. the probability distribution of z is

$$P_{\mathbf{z}}(z) = \frac{\theta^{2z}}{\theta^2 + \theta^4 + \theta^6}$$

where θ is some fixed non zero real number.

Whatever the value of θ , $P_{\mathbf{z}}(z) > 0$ for $z = 1, 2, 3$ and $P_{\mathbf{z}}(1) + P_{\mathbf{z}}(2) + P_{\mathbf{z}}(3) = 1$. Therefore z is a well-defined random variable, even if the value of θ is unknown.

We call θ a **parameter**, that is some constant, usually unknown involved in a probability distribution.

Mean and variance of a discrete r.v.

Mean (or expected value) of a discrete random variable z is defined as

$$E[\mathbf{z}] = \mu = \sum_{z \in \mathcal{Z}} z P_{\mathbf{z}}(z)$$

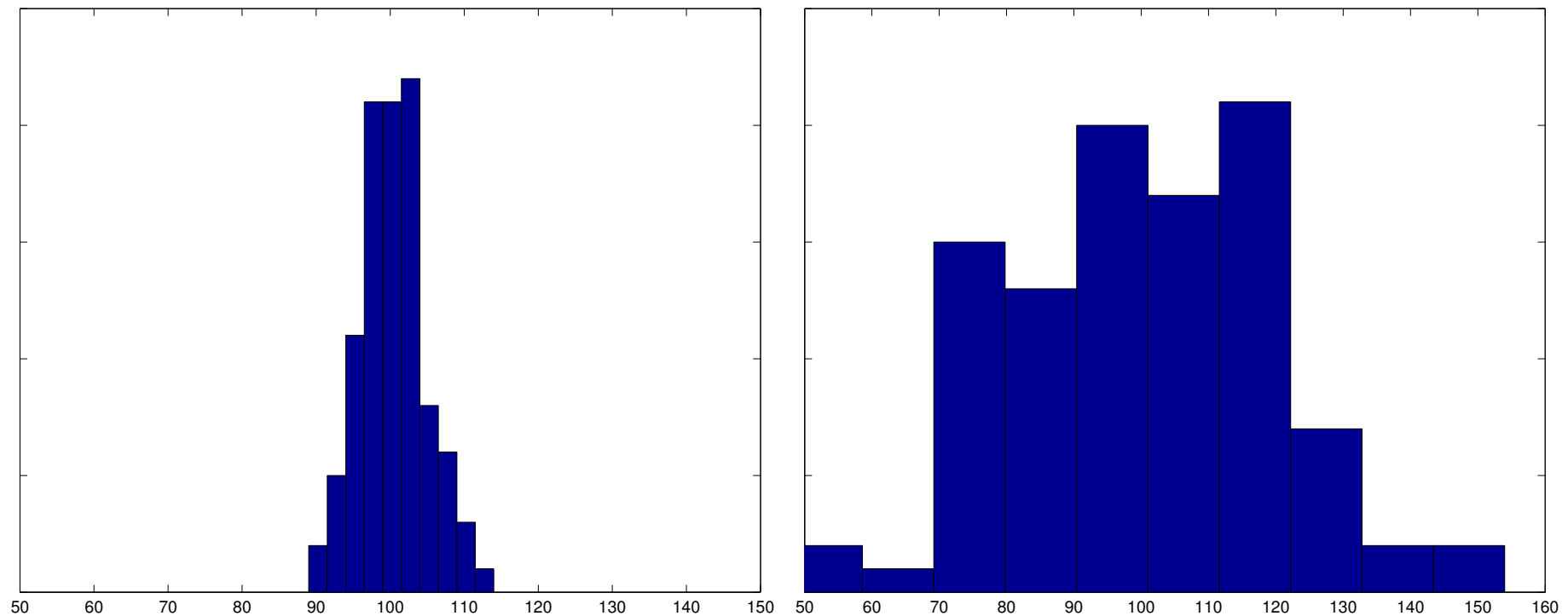
- The word “average” is not a synonym of the word “mean”.
- The mean is not necessarily a value that belongs to \mathcal{Z} .

Variance of a discrete random variable z is defined as

$$\text{Var} [\mathbf{z}] = \sigma^2 = \sum_{z \in \mathcal{Z}} (z - E[\mathbf{z}])^2 P_{\mathbf{z}}(z) \equiv E[\mathbf{z}^2] - (E[\mathbf{z}])^2$$

- The variance is a measure of the dispersion of the probability distribution of the random variable around its mean.

Probability distributions



Two discrete r.v. distributions having the same mean but different variance.

Std. deviation and moments of a discrete r.v.

Standard deviation of a discrete random variable \mathbf{z} is defined as the positive square root of the variance.

$$\text{Std} [\mathbf{z}] = \sqrt{\text{Var} [\mathbf{z}]} = \sigma$$

Moment: for any positive integer r , the r th moment of the probability distribution about the mean is

$$\mu_r = E[\mathbf{z}^r] = \sum_z (z - E[\mathbf{z}])^r P_{\mathbf{z}}(z)$$

Skewness of a discrete random variable \mathbf{z} is defined as

$$\gamma = \frac{\mu_3}{\sigma^3}$$

- Distributions with positive skewness have long tails to the right, and distributions with negative skewness have long tails to the left.

Linear combinations

- The expectation value of a linear combination of r.v.'s is simply the linear combination of their respective expectation values

$$E[ax + by] = aE[x] + bE[y]$$

i.e., expectation is a linear statistic.

- Since the variance is not a linear statistic, we have

$$\begin{aligned}\text{Var}[ax + by] &= a^2\text{Var}[x] + b^2\text{Var}[y] + 2ab(E[xy] - E[x]E[y]) = \\ &= a^2\text{Var}[x] + b^2\text{Var}[y] + 2ab\text{Cov}[x, y]\end{aligned}$$

where $\text{Cov}[x, y]$ is called **covariance**.

Correlation

- The correlation coefficient is

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\text{Cov}[\mathbf{x}, \mathbf{y}]}{\sqrt{\text{Var}[\mathbf{x}] \text{Var}[\mathbf{y}]}}$$

It is easily shown that $-1 \leq \rho(\mathbf{x}, \mathbf{y}) \leq 1$.

- If \mathbf{x} and \mathbf{y} are two independent random variables then $\text{Cov}[\mathbf{x}, \mathbf{y}] = 0$.
- In general, if the random variables $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ are independent, then

$$\text{Var} \left[\sum_{i=1}^k c_i \mathbf{x}_i \right] = \sum_{i=1}^k c_i^2 \sigma_i^2$$

Entropy

Given a discrete r.v. \mathbf{z} , the **entropy** of its probability distribution is defined by

$$H(P) = - \sum_{z \in \mathcal{Z}} P_{\mathbf{z}}(z) \log P_{\mathbf{z}}(z)$$

- H is a measure of the unpredictability of a r.v.
- If there are s possible values for a r.v. \mathbf{z} , the entropy is maximized if $P_{\mathbf{z}}(z) = 1/s$ and takes the value $\log s$.
- The entropy is minimized if only one value of \mathbf{z} is possible.
- Since H depends only on the probabilities of the various values of \mathbf{z} and not on the actual values themselves, it can be thought as a function of the probability distribution P rather than of \mathbf{z} .
- Although entropy and variance measure in some sense the uncertainty of a r.v., the entropy has a different interpretation since it depends only on the probabilities of the different values and not on the values themselves.

Relative entropy

Let us consider two different discrete probability distributions on the same set of values

$$P_0 = P_{\mathbf{z}_0}(z), \quad P_1 = P_{\mathbf{z}_1}(z)$$

where $P_0(z) > 0$ if and only if $P_1(z) > 0$.

The **relative entropies** associated with these two distributions are

$$H(P_0||P_1) = \sum_z P_0(z) \log \frac{P_0(z)}{P_1(z)} \quad H(P_1||P_0) = \sum_z P_1(z) \log \frac{P_1(z)}{P_0(z)}$$

These quantities measure the dissimilarity between the two distributions. In order to satisfy the reciprocity requirement the **divergence quantity**

$$J(P_0, P_1) = H(P_0||P_1) + H(P_1||P_0)$$

is typically used.

Marginal probability

- Consider a probabilistic model described by n discrete random variables.
- A fully specified probabilistic model gives the *joint probability* for every combination of the values of the n r.v.s.
- The model is specified by the values of the probabilities

$$\text{Prob} \{ \mathbf{z}_1 = z_1, \mathbf{z}_2 = z_2, \dots, \mathbf{z}_n = z_n \} = P(z_1, z_2, \dots, z_n)$$

for every possible assignment of values z_1, \dots, z_n to the variables.

- *Marginal probabilities* for subsets of the n variables can be found by summing over all possible combinations of values for the other variables.

$$P(z_1, \dots, z_m) = \sum_{\tilde{z}_{m+1}} \cdots \sum_{\tilde{z}_n} P(z_1, \dots, z_m, \tilde{z}_{m+1}, \dots, \tilde{z}_n)$$

Conditional probability

- *Conditional probabilities* for one subset of variables $\{x_i : i \in S_1\}$ given values for another disjoint subset $\{x_j : j \in S_2\}$ where $S_1 \cap S_2 = \emptyset$, are defined as ratios of marginal probabilities

$$P(\{x_i : i \in S_1\} | \{x_j : j \in S_2\}) = \frac{P(\{x_i : i \in S_1\}, \{x_j : j \in S_2\})}{P(\{x_j : j \in S_2\})}$$

Marginal/conditional: example

Consider a probabilistic model of the day's weather based on $n = 3$ discrete variables, where

1. z_1 represents the sky condition and takes value in the finite set {CLEAR, CLOUDY}.
2. z_2 represents the barometer trend and takes value in the finite set {RISING, FALLING},
3. z_3 represents the humidity in the afternoon and takes value in {DRY, WET}.

Marginal/conditional: example (II)

Let the joint distribution be given by the table

z_1	z_2	z_3	$P(z_1, z_2, z_3)$
CLEAR	RISING	DRY	0.4
CLEAR	RISING	WET	0.07
CLEAR	FALLING	DRY	0.08
CLEAR	FALLING	WET	0.10
CLOUDY	RISING	DRY	0.09
CLOUDY	RISING	WET	0.11
CLOUDY	FALLING	DRY	0.03
CLOUDY	FALLING	DRY	0.12

From the joint distribution we can calculate the marginal probabilities $P(CLEAR, RISING) = 0.47$ and $P(CLOUDY) = 0.35$ and the conditional value

$$P(DRY|CLEAR, RISING) =$$

$$= \frac{P(DRY, CLEAR, RISING)}{P(CLEAR, RISING)} = \frac{0.40}{0.47} \approx 0.85$$

Bernoulli trial

- A **Bernoulli trial** is a single trial with two possible outcomes, often called “success” and “failure”.
- The probability of success is denoted by p and the probability of failure by $(1 - p)$.
- A **Bernoulli random variable z** is a discrete r.v. associated with the Bernoulli trial. It takes $z = 0$ with probability $(1 - p)$ and $z = 1$ with probability p .
- The probability distribution of z can be written in the form

$$\text{Prob} \{z = z\} = P_z(z) = p^z(1 - p)^{1-z} \quad z = 0, 1$$

- Note that $E[z] = p$ and $\text{Var}[z] = p(1 - p)$.

The Binomial distribution

- A **binomial random variable** is the number of successes in a *fixed number* N of *independent Bernoulli* trials with the *same probability* of success for each trial.
- Example: the number z of heads in N tosses of a coin.
- The probability distribution is given by

$$\text{Prob} \{ \mathbf{z} = z \} = P_{\mathbf{z}}(z) = \binom{N}{z} p^z (1 - p)^{N-z} \quad z = 0, 1, \dots, N$$

- The mean of the distribution is $\mu = Np$.
- The Bernoulli distribution is a special case ($N = 1$) of the binomial distribution.
- For small p , the probability of having at least 1 success in N trials is proportional to N , as long as Np is small.

The Geometric distribution

- A r.v. z has a **geometric distribution** if it represents the number of successes before the first failure in a sequence of independent Bernoulli trials with probability of success p .
- Its probability distribution is

$$P_{\mathbf{z}}(z) = (1 - p)p^z \quad z = 0, 1, 2, \dots$$

- A r.v. z has a **generalized geometric distribution** if it represents the number of Bernoulli trials preceding but not including the $k + 1$ th failure. Its distribution is

$$P_{\mathbf{z}}(z) = \binom{z}{k} p^{z-k} (1 - p^{k+1}), \quad z = k, k + 1, k + 2, \dots$$

The Poisson distribution

- A r.v. z has a **Poisson distribution** with parameter $\lambda > 0$ is

$$P_{\mathbf{z}}(z) = \frac{e^{-\lambda} \lambda^z}{z!}$$

- The Poisson distribution is a limiting form of the binomial distribution.
- If the number of trials N is large, the probability of success p of each trial is small and the product $Np = \lambda$ is moderate, then the binomial distribution of z successes is very close to the probability that a Poisson random variable with parameter λ takes the value z .
- $\text{Prob} \{z \geq 1\} = 1 - \text{Prob} \{z = 0\} = 1 - e^{-\lambda}$.
- $E[z] = \lambda, \text{Var} [z] = \lambda$.

Mean and variances of the distributions

Distribution	Mean	Variance
Bernoulli	p	$p(1 - p)$
Binomial	Np	$Np(1 - p)$
Geometric	$p/(1 - p)$	$p/(1 - p)^2$
Poisson	λ	λ

Continuous random variable

- **Continuous random variables** take their value in some continuous range of values. Consider a real random variable \mathbf{z} whose range is the set of real numbers. The following quantities can be defined:

Definition 1. *The distribution function of \mathbf{z} is the function*

$$F_{\mathbf{z}}(z) = \text{Prob} \{ \mathbf{z} \leq z \} \quad (1)$$

Definition 2. *The density function of a real random variable \mathbf{z} is the derivative of the distribution function:*

$$p_{\mathbf{z}}(z) = \frac{dF_{\mathbf{z}}(z)}{dz} \quad (2)$$

- Probabilities of continuous r.v. are not allocated to specific values but rather to interval of values. Specifically

$$\text{Prob} \{ a < z < b \} = \int_a^b p_{\mathbf{z}}(z) dz$$

Mean, variance,... of a continuous r.v.

Consider a continuous r.v. having range (l, h) and density function $p(z)$. We can define

Mean:

$$\mu = \int_l^h zp(z)dz$$

Variance:

$$\sigma^2 = \int_l^h (z - \mu)^2 p(z) dz$$

Other quantities of interest are the *moments* :

$$\mu_r = E[z^r] = \int_l^h z^r p(z) dz$$

The moment of order $r = 1$ is the *mean* of z .

The Chebyshev's inequality

- Let z be a **generic** random variable, discrete or continuous , having mean μ and variance σ^2 .
- The Chebyshev's inequality states that for any positive constant d

$$\text{Prob} \{ |z - \mu| \geq d \} \leq \frac{\sigma^2}{d^2}$$

- An experimental validation of the Chebyshev's inequality can be found in the MATLAB file `cheby.m`.

Bivariate probability distribution

- Let us consider two r.v.'s x and y and their joint density function $p_{\mathbf{x},\mathbf{y}}(x, y)$.
- We define **marginal density** the quantity

$$m(x) = \int_{-\infty}^{\infty} p_{\mathbf{x},\mathbf{y}}(x, y) dy$$

- We define **conditional density** the quantity

$$p(y|x) = \frac{p(x, y)}{m(x)}$$

which is, in loose terms, the probability that y is in dy about y assuming $x = x$.

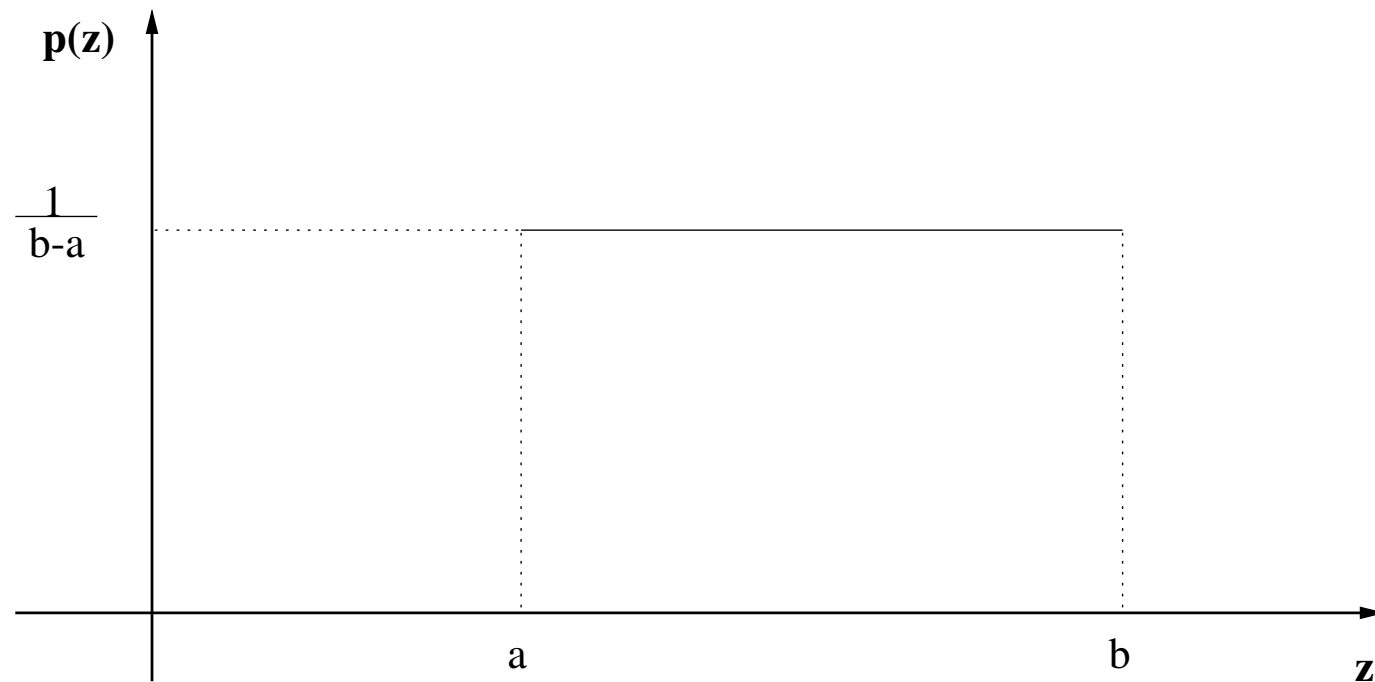
- If x and y are independent

$$p_{\mathbf{x},\mathbf{y}}(x, y) = p_{\mathbf{x}}(x)p_{\mathbf{y}}(y), \quad m(x) = p_{\mathbf{x}}(x), \quad p(y|x) = p_{\mathbf{y}}(y)$$

Uniform distribution

A random variable z is said to be **uniformly distributed** on the interval (a, b) (also $z \sim \mathcal{U}(a, b)$) if its probability density function is given by

$$p(z) = \begin{cases} \frac{1}{b-a} & \text{if } a < z < b \\ 0, & \text{otherwise} \end{cases}$$



Exponential distribution

A continuous random variable z is said to be **exponentially distributed** with rate $\lambda > 0$ (also $z \sim \mathcal{E}(\lambda)$) if its probability density function is given by

$$p_{\mathbf{z}}(z) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

- The mean of z is $1/\lambda$.
- The variance of z is $1/\lambda^2$.
- It can be considered as continuous approximation to the geometric distribution.
- Like the geometric distribution it satisfies the **memoryless property**

$$\text{Prob} \{z \geq z_1 + z_2 | z = z_1\} = \text{Prob} \{z \geq z_2\}$$

- It is used to describe physical phenomena (e.g. radioactive decay time or failure time).

The Gamma distribution

We say that \mathbf{z} has a **gamma distribution** with parameters (n, λ) if its density function is

$$p_{\mathbf{z}}(z) = \begin{cases} \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!} & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$$

- $E[\mathbf{z}] = n\lambda^{-1}$, $\text{Var}[\mathbf{z}] = n\lambda^{-2}$.
- It is the distribution of the sum of n i.i.d. r.v. having exponential distribution with rate λ .
- The exponential distribution is a special case ($n = 1$) of the gamma distribution.

Normal distribution: the scalar case

A continuous scalar random variable z is said to be **normally distributed** with parameters μ and σ^2 (also $z \sim \mathcal{N}(\mu, \sigma^2)$) if its probability density function is given by

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-\mu)^2}{2\sigma^2}}$$

- The mean of z is μ ; the variance of z is σ^2 .
- The coefficient in front of the exponential ensures that $\int p(z)dz = 1$.
- The probability that an observation z from a normal r.v. is within 2 standard deviations from the mean is 0.95.
- If $\mu = 0$ and $\sigma^2 = 1$ the distribution is defined **standard normal**. We will denote its distribution function $F_z(z) = \Phi(z)$.
- Given a normal r.v. $x \sim \mathcal{N}(\mu, \sigma^2)$, the r.v. $z = (x - \mu)/\sigma$ has a standard normal distribution.

Important relations

$$\mathbf{z} \in \mathcal{N}(\mu, \sigma^2)$$

$$\text{Prob} \{ \mu - \sigma \leq \mathbf{z} \leq \mu + \sigma \} = 0.683$$

$$\text{Prob} \{ \mu - 1.282\sigma \leq \mathbf{z} \leq \mu + 1.282\sigma \} = 0.8$$

$$\text{Prob} \{ \mu - 1.645\sigma \leq \mathbf{z} \leq \mu + 1.645\sigma \} = 0.9$$

$$\text{Prob} \{ \mu - 1.96\sigma \leq \mathbf{z} \leq \mu + 1.96\sigma \} = 0.95$$

$$\text{Prob} \{ \mu - 2\sigma \leq \mathbf{z} \leq \mu + 2\sigma \} = 0.954$$

$$\text{Prob} \{ \mu - 3\sigma \leq \mathbf{z} \leq \mu + 3\sigma \} = 0.997$$

Normal distribution: the multivariate case

Let \mathbf{z} be a random vector ($n \times 1$).

The vector is said to be **normally distributed** with parameters $\mu_{(n \times 1)}$ and $\Sigma_{(n \times n)}$ (also $\mathbf{z} \sim \mathcal{N}(\mu, \Sigma)$) if its probability density function is given by

$$p_{\mathbf{z}}(z) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\Sigma)}} \exp \left\{ -\frac{1}{2} (z - \mu)^T \Sigma^{-1} (z - \mu) \right\}$$

It follows that

- the mean $E[\mathbf{z}] = \mu$ is a n -dimensional vector,
- the matrix $\Sigma = E[(\mathbf{z} - \mu)(\mathbf{z} - \mu)^T]$ is the covariance matrix. This matrix is symmetric and therefore has $n(n + 1)/2$ parameters.

Normal multivariate distribution (II)

The quantity

$$\Delta = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

which appears in the exponent of $p_{\mathbf{z}}$ is called the *Mahalanobis distance* from z to μ .

It can be shown that

- the surfaces of constant probability density are hyperellipsoids on which Δ^2 is constant;
- the *principal axes* are given by the eigenvectors $u_i, i = 1, \dots, n$ of Σ which satisfy

$$\Sigma u_i = \lambda_i u_i \quad i = 1, \dots, n$$

where λ_i are the corresponding eigenvalues.

- the eigenvalues λ_i give the variances along the principal directions.

Normal multivariate distribution (III)

If the covariance matrix Σ is diagonal then

- the contours of constant density are hyperellipsoids with the principal directions aligned with the coordinate axes.
- the components of \mathbf{z} are then *statistically independent* since the distribution of \mathbf{z} can be written as the product of the distributions for each of the components separately in the form

$$p_{\mathbf{z}}(\mathbf{z}) = \prod_{i=1}^n p(z_i)$$

- the total number of independent parameters in the distribution is $2n$.
- if $\sigma_i = \sigma$ for all i , the contours of constant density are hyperspheres.

Bivariate normal distribution

Consider a bivariate normal density whose mean is $\mu = [\mu_1, \mu_2]^T$ and the covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

The correlation coefficient is

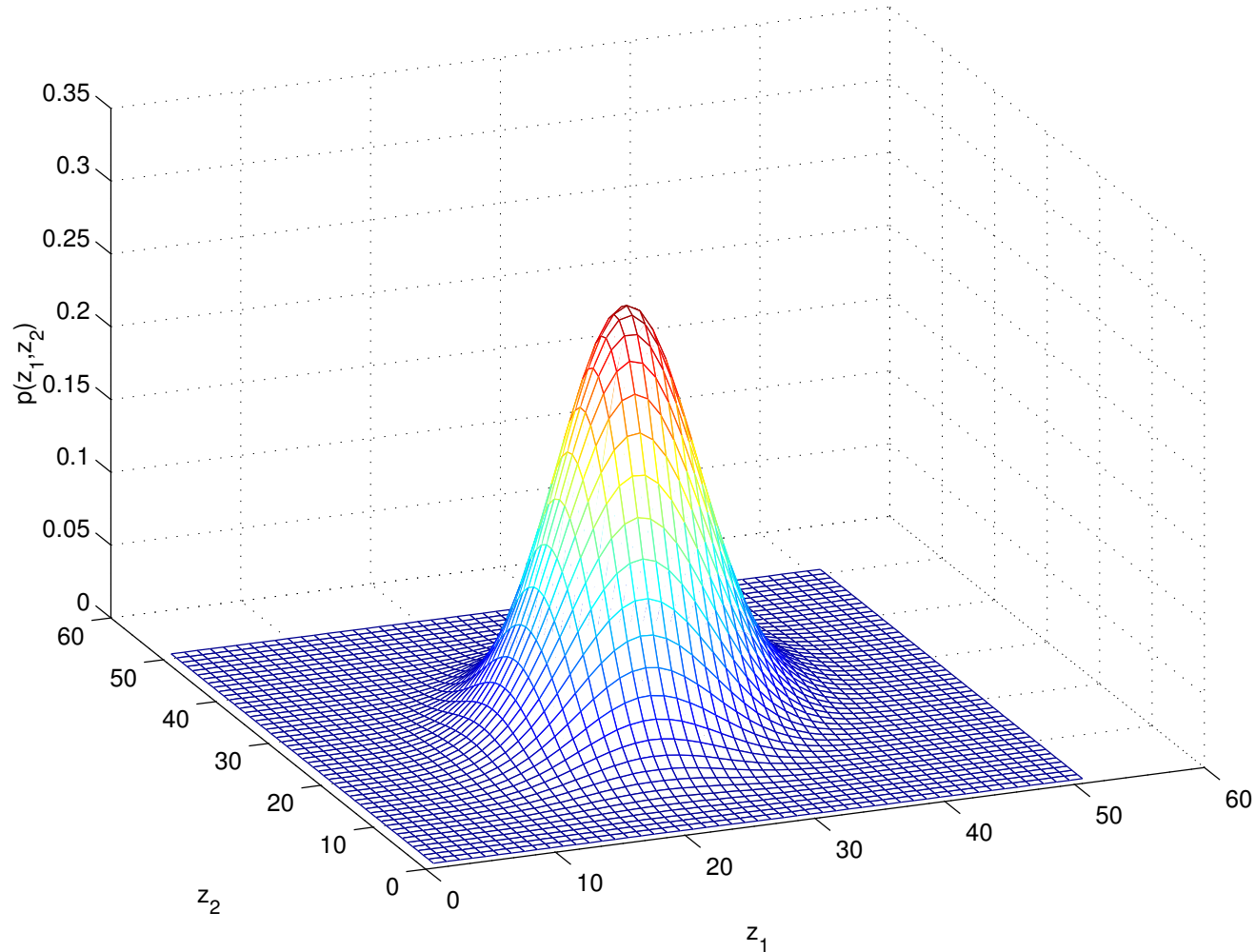
$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

It can be shown that the general bivariate normal density has the form

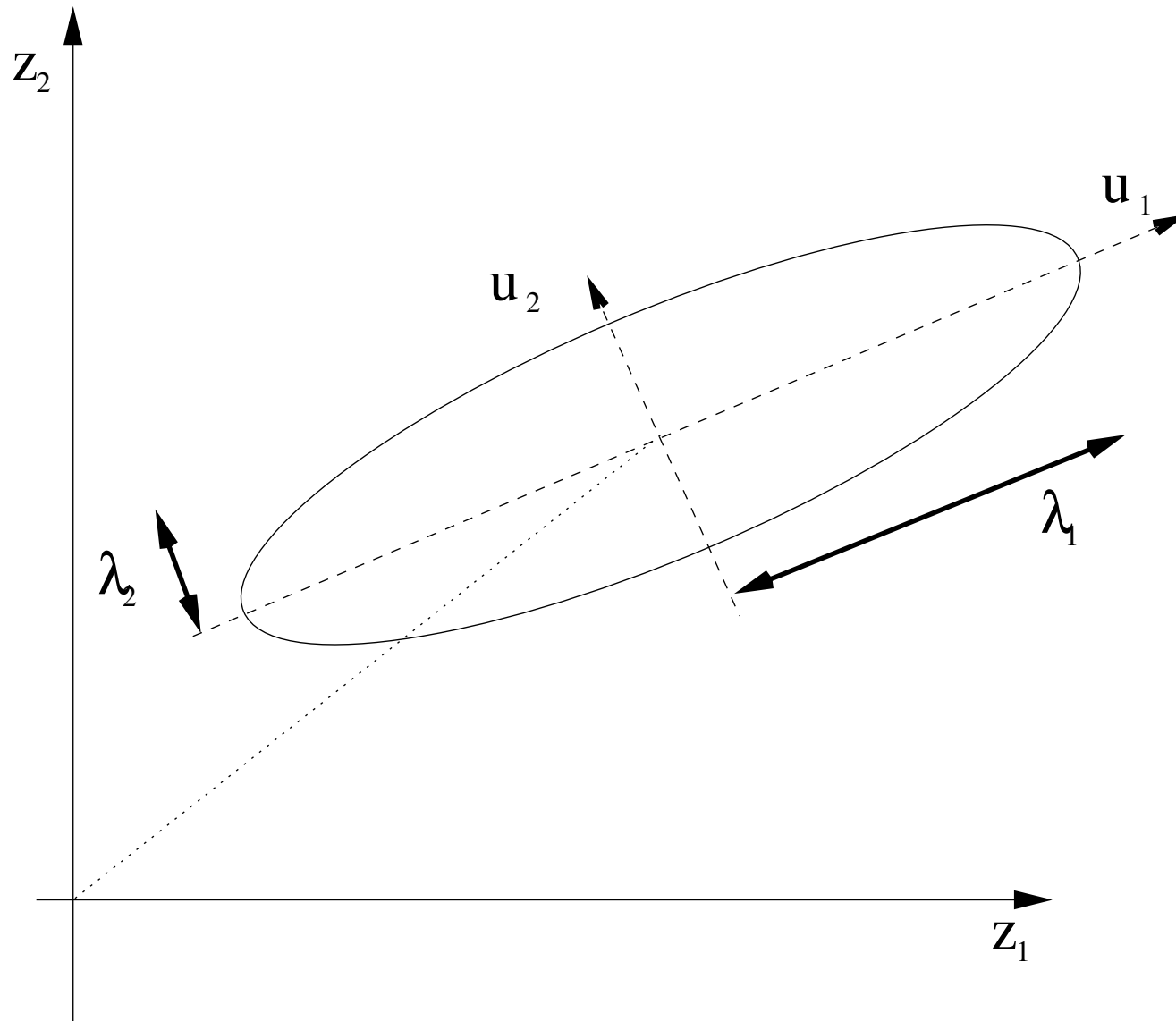
$$p(z_1, z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left[\left(\frac{z_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{z_1 - \mu_1}{\sigma_1} \right) \left(\frac{z_2 - \mu_2}{\sigma_2} \right) + \left(\frac{z_2 - \mu_2}{\sigma_2} \right)^2 \right] \right]$$

Bivariate normal distribution

Let $\Sigma = [1.2919, 0.4546; 0.4546, 1.7081]$.



Bivariate normal distribution (prj)



Marginal and conditional distributions

One of the important properties of the multivariate normal density is that all conditional and marginal probabilities are also normal. Using the relation

$$p(z_2|z_1) = \frac{p(z_1, z_2)}{p(z_1)}$$

we find that $p(z_2|z_1)$ is a normal distribution $\mathcal{N}(\mu_{2|1}, \sigma_{2|1}^2)$, where

$$\mu_{2|1} = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (z_1 - \mu_1)$$

$$\sigma_{2|1}^2 = \sigma_2^2 (1 - \rho^2)$$

Note that

- $\mu_{2|1}$ is a linear function of z_1 : if the correlation coefficient ρ is positive, the larger z_1 , the larger $\mu_{2|1}$.
- if there is no correlation between z_1 and z_2 , we can ignore the value of z_1 and use always μ_2 to estimate z_2 .

Upper critical points

Definition 3. The upper critical point of a continuous r.v. \mathbf{x} is the smallest number x_α such that

$$1 - \alpha = \text{Prob} \{ \mathbf{x} \leq x_\alpha \} = F(x_\alpha)$$

- We will denote with z_α the upper critical points of the standard normal density.

$$1 - \alpha = \text{Prob} \{ \mathbf{z} \leq z_\alpha \} = F(z_\alpha) = \Phi(z_\alpha)$$

- Note that the following relations hold

$$z_\alpha = \Phi^{-1}(1 - \alpha), \quad z_{1-\alpha} = -z_\alpha, \quad 1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-z^2/2} dz$$

- Here we list the most commonly used values of z_α .

α	0.1	0.075	0.05	0.025	0.01	0.005	0.001	0.0005
z_α	1.282	1.440	1.645	1.967	2.326	2.576	3.090	3.291

The sum of i.i.d. random variables

Suppose that $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ are i.i.d. (identically and independently distributed) random variables, discrete or continuous, each having a probability distribution with mean μ and variance σ^2 . Let us consider the two derived r.v., that is the sum

$$\mathbf{S}_N = \mathbf{z}_1 + \mathbf{z}_2 + \dots + \mathbf{z}_N$$

and the average

$$\bar{\mathbf{z}} = \frac{\mathbf{z}_1 + \mathbf{z}_2 + \dots + \mathbf{z}_N}{N}$$

The following relations hold

$$E[\mathbf{S}_N] = N\mu, \quad \text{Var}[\mathbf{S}_N] = N\sigma^2$$

$$E[\bar{\mathbf{z}}] = \mu, \quad \text{Var}[\bar{\mathbf{z}}] = \frac{\sigma^2}{N}$$

See the MATLAB script `sum_rv.m`.

The central limit theorem

Theorem 1. Assume that $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$ are i.i.d. random variables, discrete or continuous, each having a probability distribution with finite mean μ and finite variance σ^2 . As $N \rightarrow \infty$, the standardized random variable

$$\frac{(\bar{\mathbf{z}} - \mu)\sqrt{N}}{\sigma}$$

which is identical to

$$\frac{(\mathbf{S}_N - N\mu)}{\sqrt{N}\sigma}$$

converges in distribution to a r.v. having the standardized **normal distribution** $\mathcal{N}(0, 1)$.

- This result holds regardless of the common distribution of \mathbf{z}_i .
- This theorem justifies the importance of the normal distribution, since many r.v. of interest are either sums or averages.

The chi-squared distribution

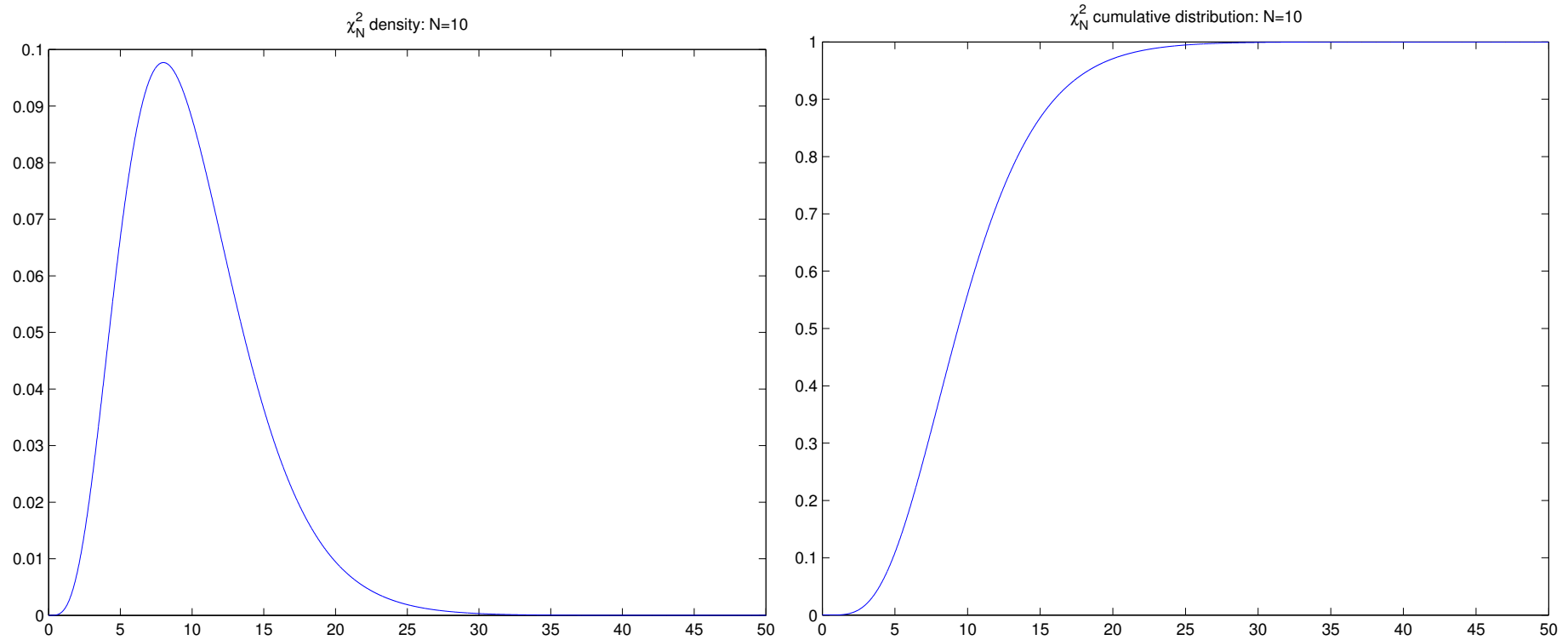
For a N positive integer, a r.v. z has a χ_N^2 distribution if

$$z = x_1^2 + \cdots + x_N^2$$

where x_1, x_2, \dots, x_N are i.i.d. random variables $\mathcal{N}(0, 1)$.

- The probability distribution is a gamma distribution with parameters $(\frac{1}{2}N, \frac{1}{2})$.
- $E[z] = N$ and $\text{Var}[z] = 2N$.
- The distribution is called “a chi-squared distribution with N degrees of freedom”.

The chi-squared distribution (II)



MATLAB script `chisq.m`.

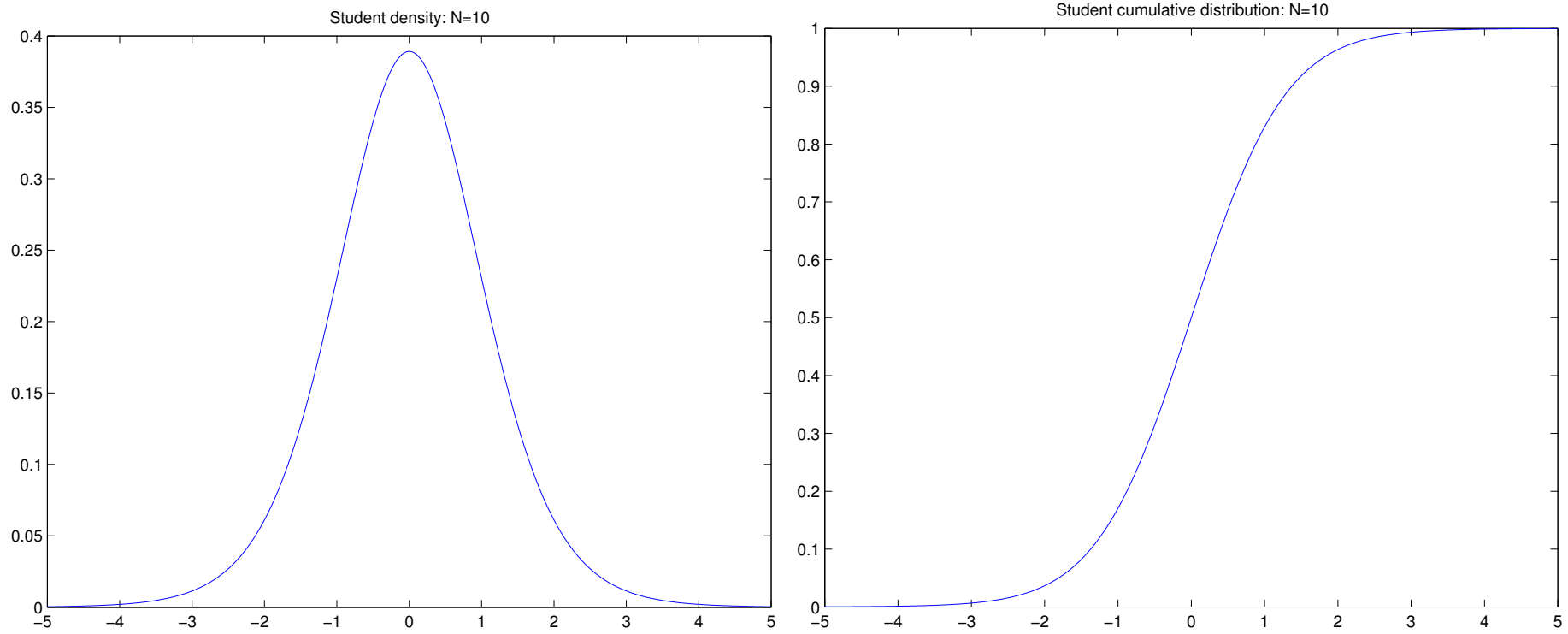
Student's t -distribution

If $\mathbf{x} \sim \mathcal{N}(0, 1)$ and $\mathbf{y} \sim \chi_N^2$ are independent then the **Student's t -distribution** with N degrees of freedom is the distribution of the r.v.

$$\mathbf{z} = \frac{\mathbf{x}}{\sqrt{\mathbf{y}/N}}$$

We denote this with $\mathbf{z} \sim t_N$.

Student's t -distribution



MATLAB script `s_stu.m`.

Notation

- In order to clarify the distinction between random variables and their values, we will use the **boldface notation for denoting a random variable** (e.g. \mathbf{z}) and the normal face notation for the eventually observed value (e.g. $z = 11$).
- The notation $P_{\mathbf{z}}(z)$ denotes the probability that the random variable \mathbf{z} take the value z . The suffix indicates that the probability relates to the random variable \mathbf{z} . This is necessary since we often discuss probabilities associated with several random variables simultaneously.
- Example: \mathbf{z} could be the age of a student before asking and $z = 22$ could be his value after the observation.

Notation (II)

- In general terms, we will denote as the **probability distribution of a random variable z** any complete description of the probabilistic behavior of z .
- For example, if z is continuous, the density function $p(z)$ or the distribution function could be examples of probability distribution.
- Given a probability distribution $F_z(z)$ the notation

$$F \rightarrow \{z_1, z_2, \dots, z_N\}$$

means that the dataset $D_N = \{z_1, z_2, \dots, z_N\}$ is a i.i.d. random sampled observed from the probability distribution F .