STAT2201

Analysis of Engineering & Scientific Data

Unit 4

Slava Vaisman

The University of Queensland
School of Mathematics and Physics

Joint Distributions (1)

Many random experiments involve not just one, but multiple random variables.

Examples.

1. We randomly select a person from a large population and measure his/her weight X and height Y.
2. We shoot at a two-dimensional target. Let X and Y be the coordinates of the point of impact.
3. We randomly select 20 people from a large population and ask their preference for a political party. Number the people from 1 to 20, and let $X_1, \dots, X_{20}$ be the measurements.

Joint Distributions (2)

- How can we specify a model for the experiments above?
- We cannot just specify the pdf or pmf of the individual random variables.
- We also need to specify the “interaction;; between the random variables. E.g., in Example 1, if the height Y is large, we expect that X is large as well.
- We need to specify the joint distribution of all the random variables $X_1, \ldots, X_n$ involved in the experiment.
- Alternatively, we need to specify the distribution of the random vector $X := (X_1, \ldots, X_n)$.
- We first show how this works for pairs of random variables.

Joint pmf

We will write $\mathbb{P}(X = x, Y = y)$ for the probability of the event $\{X = x\} \cap \{Y = y\}$.

Definition

Let (X, Y) be a discrete random vector. The function $(x, y) \to \mathbb{P}(X = x, Y = y)$ is called the *joint probability mass function* of X and Y.

Example.

- In a box are three dice. Die 1 is a normal die; die 2 has no 6 face, but instead two 5 faces; die 3 has no 5 face, but instead two 6 faces.
- The experiment consists of selecting a die at random, followed by a toss with that die.
- Let X be the die number that is selected, and let Y be the face value of that die. The joint pmf of X and Y is specified over.

Joint and Marginal pmfs

x \ y	1	2	3	4	5	6	$\sum$
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	0	$\frac{1}{3}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	0	$\frac{1}{9}$	$\frac{1}{3}$
$\sum$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

The pmf's of X and Y , the so-called marginal pmf's, can be found by summing up over respectively the y's and the x's, e.g.,

$$\mathbb{P}(X = x) = \sum_{y} \mathbb{P}(X = x, Y = y).$$

Joint pmf

Let $\Omega_{X,Y}$ be the set of possible outcomes of (X, Y). We have for all $B \subset \Omega_{X,Y}$,

$$\mathbb{P}((X, Y) \in B) = \sum_{(x,y)\in B} \mathbb{P}(X = x, Y = y).$$

- An important way of creating joint pmfs is by starting with the marginal pmfs of X and Y and then to define the events $\{X = x\}$ and $\{Y = y\}$ to be independent, for all x and y.
- We then have (definition of independent events)

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y).$$

Example.

- Repeat the experiment above with three normal dice. Since the events $\{X = x\}$ and $\{Y = y\}$ should be independent, each entry in the pmf table is $1/3 \times 1/6$.
- Clearly in the first experiment not all events $\{X = x\}$ and $\{Y = y\}$ are independent (why not?).

Joint pdf

The analogue of the pmf for continuous X and Y is the joint pdf.

Definition

We say that the random variables X and Y have a joint probability density function f if, for all events $\{(X, Y) \in A\}$, where A is a subset of $\mathbb{R}^2$ (the plane), we have

$$\mathbb{P}((X, Y) \in A) = \int \int_A f(x, y) \, \mathrm{d}x \, \mathrm{d}y.$$

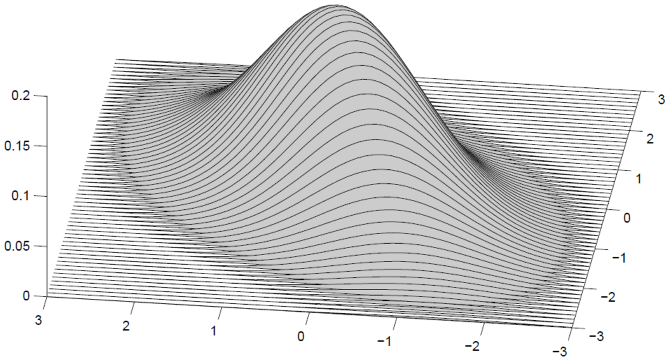
- We often write $f_{X,Y}$ for f.
- Note that the calculation of probabilities has been reduced to integration over a set A.

Bivariate normal distribution

We say that Z_1 and Z_2 have a bi-variate Gaussian (or normal) distribution with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ if the joint density function is given by

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

Bivariate normal distribution



$\rho = 0.5$

Properties (1)

The joint pdf has the following properties (similar to the 1-dimensional case):

1. $f(x, y) \geq 0$, for all x and y.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1$.

- Any function f satisfying these conditions, and for which the integral is well-defined, can be a joint pdf.
- $f(x, y)$ can be interpreted as the "infinitesimal" probability that $X = x$ and $Y = y$:

$$\mathbb{P}(x \leq X \leq x + h, y \leq Y \leq y + h) =$$
$$= \int_{x}^{x+h} \int_{y}^{y+h} f(u, v)\, \mathrm{d}u\, \mathrm{d}v \approx h^2 f(x, y).$$

Properties (2)

- We can obtain the marginal pdf of X, say f_X, by integrating f over all y:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y.$$

- Similar for Y.
- Finally, we can define the joint cdf.
- Let X and Y be two random variables observed from the same random experiment.
- We define the joint cumulative distribution function as $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$.
- If X and Y are continuous random variables, the joint pdf is then

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

Independence

- Independence. Let X and Y be two random variables.

Definition

We say X and Y are independent random variables if any event defined by X is independent of every event defined by Y.

- That is, if X and Y are independent, we can always say that

  $$\mathbb{P}(\{a < X \leq b\} \cap \{c < Y \leq d\}) = \mathbb{P}(a < X \leq b)\, \mathbb{P}(c < Y \leq d)$$

  for any possible choice of a, b, c and d.
- It follows that X and Y are independent if and only if

$$F_{X,Y}(x, y) = F_X(x)\, F_Y(y).$$

Independence

- Hence, for continuous random variables, X and Y are independent if and only if, for all x and y,

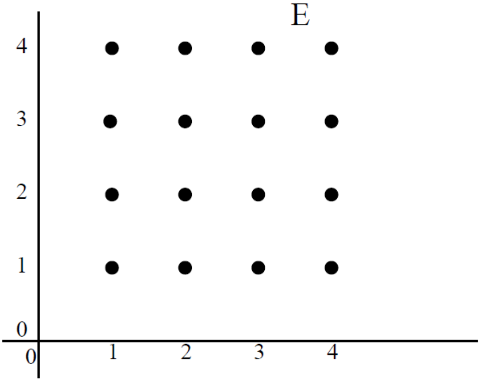
$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y).$$

- For discrete random variables, X and Y are independent if and only if, for all x and y,

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\, \mathbb{P}(Y = y).$$

Independence

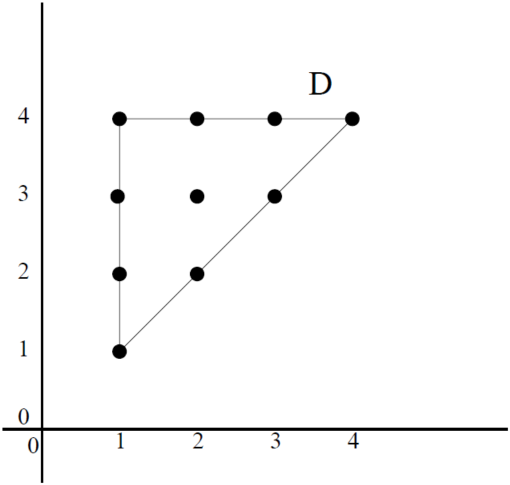
Example. We draw at random a point (X, Y) from the 16 points on the square E below.



Clearly X and Y are independent.

Conditional distributions

Example. We draw at random a point (X, Y) from the 10 points on the triangle D below.



Conditional distributions

The joint and marginal pmfs are easy to determine:

$$\mathbb{P}(X = i, Y = j) = \frac{1}{10}, \quad (i, j) \in D.$$

and,

$$\mathbb{P}(X = i) = \frac{5 - i}{10}, \quad i \in \{1, 2, 3, 4\},$$

$$\mathbb{P}(Y = j) = \frac{j}{10}, \quad j \in \{1, 2, 3, 4\}.$$

Conditional distributions

- Clearly X and Y are not independent;

$$\mathbb{P}(X = 2, Y = 2) = \frac{1}{10} \neq \frac{3}{10}\frac{2}{10} = \mathbb{P}(X = 2)\,\mathbb{P}(Y = 2).$$

- In fact, if we know that $X = 2$, then Y can only take the values $j = 2, 3$ or 4.
- The corresponding probabilities are

$$\mathbb{P}(Y = j \mid X = 2) = \frac{\mathbb{P}(X = 2, Y = j)}{\mathbb{P}(X = 2)} = \frac{1/10}{3/10} = \frac{1}{3}.$$

- We thus have determined the **conditional pmf** of Y given $X = 2$.

Conditional distributions

Definition

If X and Y are discrete and $\mathbb{P}(X = x) > 0$, then

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}$$

for all y, give the **conditional pmf** of Y given $X = x$.

We can extend this to general Y:

Definition

If X is discrete and $\mathbb{P}(X = x) > 0$, then

$$F_Y(Y = y \mid X = x) = \mathbb{P}(Y \leq y \mid X = x) = \frac{\mathbb{P}(X = x, Y \leq y)}{\mathbb{P}(X = x)},$$

gives the **conditional cdf** of Y given $X = x$.

Conditional distributions

- The corresponding density (if it exists) is the conditional pdf of Y given $X = x$, denoted

$$f_Y(y \mid x).$$

- The corresponding density is called the conditional pdf of Y given $X = x$:

$$f_Y(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

Conditional expectation

- For discrete X and Y, $\mathbb{P}(Y = y \mid X = x)$ is a genuine pmf, for each fixed x.
- Hence, we can assign to it an expectation:

$$\mathbb{E}[Y \mid X] = \sum_y y\, \mathbb{P}(Y = y \mid X = x).$$

- Similarly, in the continuous case we can define

$$\mathbb{E}[Y \mid X] = \int_y y\, f_Y(y \mid x)\, \mathrm{d}y.$$

- This number $\mathbb{E}[Y \mid X]$ is called the **conditional expectation** of Y given $X = x$.

Multiple Random Variables

- Suppose $X_1, X_2, \ldots, X_n$ are random variables pertaining to some random experiment. The joint cdf F is defined by

$$F(x_1, x_2, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n).$$

- which completely specifies the probability distribution of the vector $(X_1, X_2, \ldots, X_n)$.
- If the X_i's are discrete, it suffices to only know the joint pmf p, defined by

$$p(x_1, \ldots, x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).$$

Multiple Random Variables

- Similarly, when the X_i's are continuous the probability distribution is completely specified by the joint pdf f (if it exists):

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \cdots \partial x_n}.$$

- Integration of f over a subset A of $\mathbb{R}^n$ gives the probability that the vector $(X_1, X_2, \ldots, X_n)$ lies in A.

Independence

- Discrete random variables $X_1, X_2, \ldots, X_n$ are said to be **independent** if, for all $x_1, x_2, \ldots, x_n$,

  $$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdot \mathbb{P}(X_2 = x_2) \cdots \mathbb{P}(X_n = x_n).$$

- Similarly, for continuous random variables with a joint density function f, independence is equivalent to

  $$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n).$$

- An infinite sequence $X_1, X_2, \ldots$ of random variables is called independent if for any finite choice of parameters $i_1, i_2, \ldots, i_n$ (none of them the same), the random variables $X_{i_1}, \ldots, X_{i_n}$ are independent.

Important remarks

- More often than not, the independence of random variables is a model assumption, rather than a consequence.

- Instead of describing a random experiment via an explicit description of Ω and $\mathbb{P}$, we will usually model the experiment through one or more (independent) random variables.

Functions of random variables

Suppose $X_1, \ldots, X_n$ are the measurements on a random experiment. Often we are interested in functions of the measurement. Examples are:

1. $X_1, \ldots, X_n$ are repeated measurements of a certain quantity. Then,
$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$
is what we are really interested in.
2. $X_1, \ldots, X_n$ are the lifetimes of the components in a series system. Then, the lifetime of the system is
$$\min\{X_1, \ldots, X_n\}.$$

In general, let the random variable Z be defined as a function of several random variables: $Z = g(X_1, \ldots, X_n)$. How can we find the pmf, pdf and/or cdf of Z?

Expected value

- Similar to the 1-dimensional case, the expected value of $Z = g(X_1, \dots, X_n)$ can be evaluated in the discrete case as

$$\mathbb{E}[Z] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \dots, x_n)\, \mathbb{P}(X_1 = x_1, \dots, X_n = x_n),$$

- and, in the continuous case as

$$\mathbb{E}[Z] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

Example

In the continuous case, find the expectation of $X + Y$. (Do the discrete case yourself.)

Expected value (linearity of expectation)

Let f be the joint pdf of X and Y, then

$$\begin{aligned}
\mathbb{E}[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)\mathrm{d}x\,\mathrm{d}y = \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,y)\mathrm{d}x\,\mathrm{d}y + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x,y)\mathrm{d}x\,\mathrm{d}y = \\
&= \int_{-\infty}^{\infty}x\,f_X(x)\mathrm{d}x + \int_{-\infty}^{\infty}y\,f_Y(y)\mathrm{d}y = \mathbb{E}[X]+\mathbb{E}[Y].
\end{aligned}$$

Note that X and Y do not have to be independent.

Expected value

- This is easily generalized to the following result:
- Suppose $X_1, \dots, X_n$ are random variables measured on the same random experiment, with means $\mu_1, \dots, \mu_n$.
- Let $Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n$ where $a, b_1, \dots, b_n$ are constants.
- Then,

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n] = \\ &= a + b_1 \mathbb{E}[X_1] + \cdots b_n \mathbb{E}[X_n] = a + \mu_1 + b_1 \cdots + b_n \mu_n.\end{aligned}$$

- That is, substitute the mean for each X.
- Another important result is as follows. If $X_1, \dots, X_n$ are independent, then

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \mathbb{E}[X_1]\mathbb{E}[X_2] \cdots \mathbb{E}[X_n].$$

Correlation

Definition

The covariance of two random variables X and Y is defined as the number

$$\mathrm{cov}(X, Y) := \mathbb{E}[X - \mathbb{E}[X])\,(Y - \mathbb{E}[Y]).$$

Basically, it is a measure for the amount of linear dependency between the variables.

Closely related to this is the correlation of X and Y, defined as

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}.$$

The covariance (and correlation) of two independent random variables is 0.

Properties of Variance and Covariance

- $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.
- $\mathrm{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.
- $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$.
- $\mathrm{cov}(aX + bY, Z) = a\,\mathrm{cov}(X, Z) + b\,\mathrm{cov}(Y, Z)$.
- $\mathrm{cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X, Y)$.
- X and Y independent $\Rightarrow \mathrm{cov}(X, Y) = 0$.

As an immediate consequence, if $X_1, \ldots, X_n$ are independent, we have

$$\begin{aligned} \mathrm{Var}(Y) = \mathrm{Var}(a + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n) = \\ = b_1^2 \mathrm{Var}(X_1) + \cdots + b_n^2 \mathrm{Var}(X_n). \end{aligned}$$

Jointly Gaussian RVs

- The n-dimensional density of the random vector

$$\mathbf{X} = (X_1, \ldots, X_n)^\top$$

  (column vector), with $X_1, \ldots, X_n$ independent and standard normal, is

$$f_X(x) = (2\pi)^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2}\mathbf{x}^\top \mathbf{x}}.$$

- We consider now the function (transformation) $Z = \boldsymbol{\mu} + B\mathbf{X}$. The pdf of Z is

$$f_{\mathbf{Z}}(z) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \mathrm{e}^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})},$$

  where $\Sigma = BB^\top$.
- Z is said to have a multi-variate Gaussian (or normal) distribution with expectation vector $\boldsymbol{\mu}$ and covariance matrix Σ.

Jointly Gaussian RVs

Example. Consider the 2-dimensional case with $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$, and

$$B = \begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2 \end{pmatrix}.$$

The covariance matrix is now

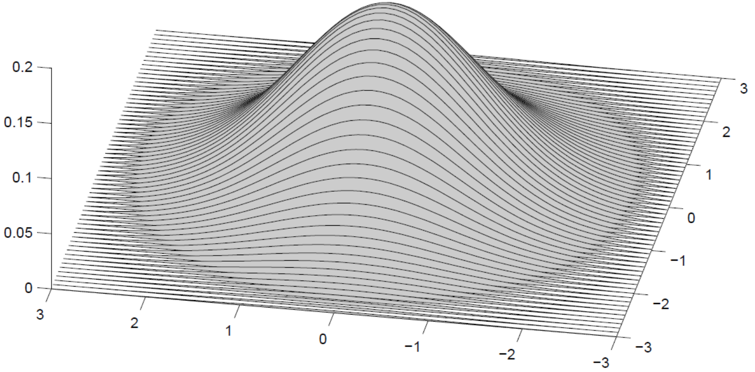
$$\Sigma = BB^\top = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Therefore, the density is

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = f_{\boldsymbol{Z}}(z_1, z_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \\ \times \left\{ \exp \frac{(z_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(z_1-\mu_1)(z_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(z_2-\mu_2)^2}{\sigma_2^2} \right\}$$

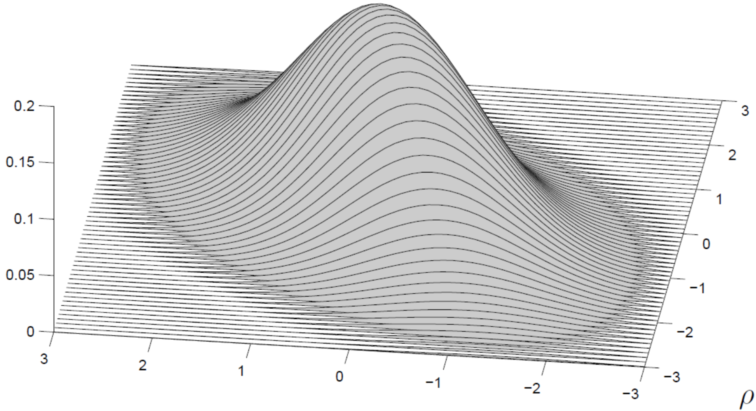
This is the pdf of the bi-variate Gaussian distribution, which we encountered earlier.

Jointly Gaussian RVs



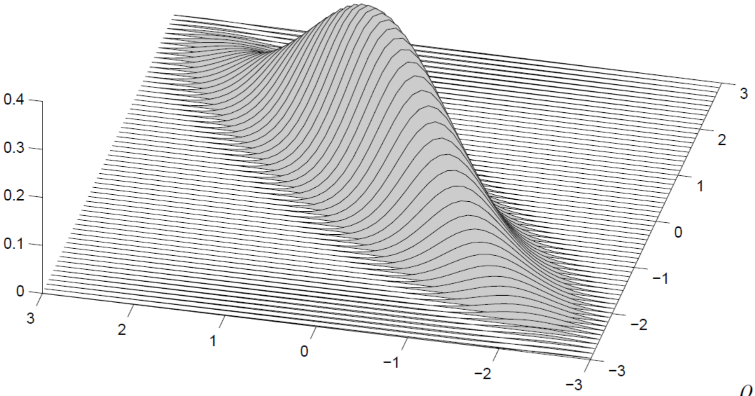
$\rho = 0.0$

Jointly Gaussian RVs



$\rho = 0.5$

Jointly Gaussian RVs



$\rho = 0.9$

Jointly Gaussian RVs

A very important property of the normal distribution is for independent

$$X_i \sim \mathrm{N}(\mu_i, \sigma_i^2), \quad i = 1, \dots, n.$$

Specifically, the random variable

$$Y = a + \sum_{i=1}^{n} b_i \, X_i,$$

is distributed

$$\mathrm{N}\left(a + \sum_{i=1}^{n} b_i \, \mu_i, \sum_{i=1}^{n} b_i^2 \, \sigma_i^2\right).$$

Jointly Gaussian RVs

Example A machine produces ball bearings with a $N(1, 0.01)$ diameter (cm). The balls are placed on a sieve with a $N(1.1, 0.04)$ diameter. The diameter of the balls and the sieve are assumed to be independent of each other. What is the probability that a ball

will fall through?

Solution

- Let $X \sim N(1, 0.01)$ and $Y \sim N(1.1, 0.04)$.
- We need to calculate $\mathbb{P}(Y > X) = \mathbb{P}(Y - X > 0)$.
- But, $T := Y - X \sim N(0.1, 0.05)$. Hence,

$$\begin{aligned}\mathbb{P}(T > 0) &= \mathbb{P}\left(\frac{Y - X}{\sqrt{0.05}} - \frac{0.1}{\sqrt{0.05}} > 0\right) \\ &= \mathbb{P}\left(Z > -\frac{0.1}{\sqrt{0.05}}\right) = 1 - \Phi(0.447) \approx 0.67.\end{aligned}$$