Probability

Lecture 4.1: Sum of RVs, Covariance and Correlation

Name teacher

TUDelft

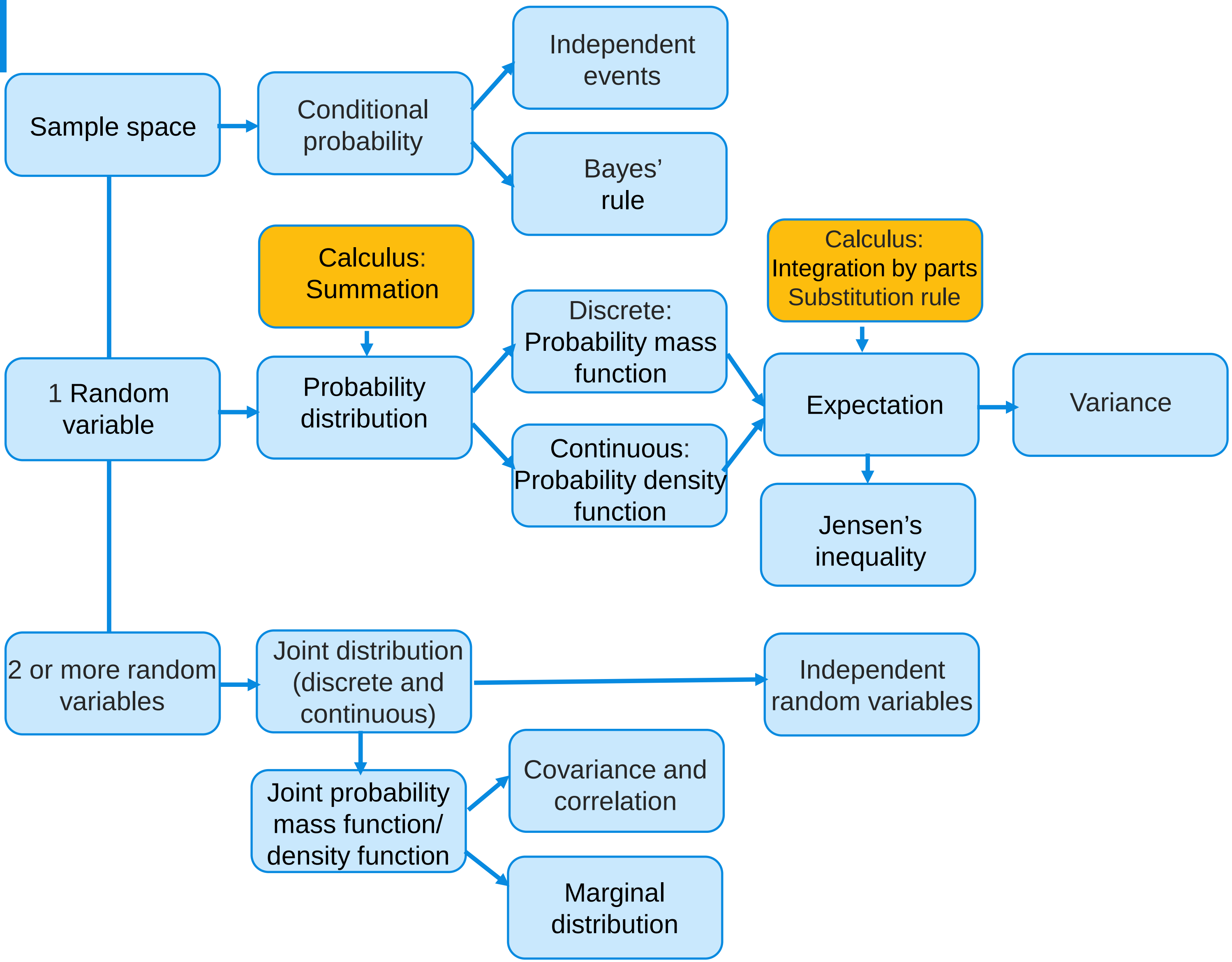
Learning objective

After this class you can

- **Check and use independence of RVs**
- **Apply the change-of-variable formula for two RVs**
- **Compute the covariance and correlation between two RVs**
- **Determine the distribution of the sum of two independent RVs**

Book: Sections 9.4, 10.1, 10.2, 10.3, 11.1, 11.2

Probability



Before this class (week 3.4 lesson 1):

Watch prelecture ‘*Independence of RVs*’

Book: Section 9.4

Programme

Expectation of a function of two RVs

Exercise

Linearity of expectations
Covariance
Independent vs uncorrelated
Correlation

Exercise

Sum of two independent RVs

Exercise

Let X, Y, Z be three independent $U(0,1)$ RVs. Compute $\mathrm{P}(X \geq YZ)$.

A) 3/4

B) 1/4

C) 1/2

D) 1/3

X and Y are two discrete RVs with joint probability mass function given by the table below. Are X and Y independent?

FeedbackFruits

		$X = a$		
		0	1	2
	-1	1/6	1/6	1/6
$Y = b$	1	0	1/2	0

A) Yes

B) No

Expectation of a function of two RV's

Definition:
Let X and Y be random variables and let $g: \mathbb{R}^2 \to \mathbb{R}$ be a function.

1. If X and Y are discrete with values $a_1, a_2, \cdots$ and $b_1, b_2, \cdots$ respectively, then
$$\mathrm{E}[g(X,Y)] = \sum_i \sum_j g(a_i, b_j)\mathrm{P}(X = a_i, Y = b_j)$$
2. If X and Y are continuous with joint probability density function f, then
$$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx dy$$

X and Y are two discrete RVs with joint probability mass function given by the table below. Compute E[X+Y].

FeedbackFruits

		$X = a$		
		0	1	2
	-1	1/6	1/6	1/6
$Y = b$	1	0	1/2	0

A) 0

B) 1/2

C) -1

D) 1

Exercise

Let X and Y be RVs such that

$$\mathrm{E}[X] = 2, \quad \mathrm{E}[Y] = 3, \quad \mathrm{Var}(X) = 4$$

(a) Compute $\mathrm{E}\left[X^2\right]$

(b) Determine the expectation of $-2X^2 + Y$

Linearity of expectations

Theorem:
For every two random variables X and Y it holds that

$$E[rX + sY + t] = rE[X] + sE[Y] + t$$

for all values r, s and t.

What about Var[X+Y]?

Covariance

Definition:
Let X and Y be two RVs. The *covariance* between X and Y is

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]$$

If Cov(X,Y)>0, then X and Y are *positively correlated*.
If Cov(X,Y)<0, then X and Y are *negatively correlated*.
If Cov(X,Y)=0, then X and Y are *uncorrelated*.

A useful way to compute the covariance is

$$\mathrm{Cov}(X,Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]$$

X and Y are two discrete RVs with joint probability mass function given by the table below. Which of the following statements is true?

		$X = a$		
		0	1	2
	-1	1/6	1/6	1/6
$Y = b$	1	0	1/2	0

A) $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$ and X and Y are dependent

B) $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$ and X and Y are independent

C) $\mathrm{E}[XY] > \mathrm{E}[X]\,\mathrm{E}[Y]$ and X and Y are dependent

D) $\mathrm{E}[XY] < \mathrm{E}[X]\,\mathrm{E}[Y]$ and X and Y are independent

Covariance

If two RVs X and Y are independent, then they are also uncorrelated.

Let X and Y be two RVs, then always

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

If X and Y are uncorrelated, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Change of units

Theorem:
Let X and Y be two RVs. Then

$$\mathrm{Cov}(rX + s, tY + u) = rt\mathrm{Cov}(X, Y)$$

for all values r, s, t and u.

Suppose the covariance of the RVs X and Y is equal to -2.5. Compute Cov(-2X+7, 5Y-2).

A) -2.5

B) -25

C) 25

D) 30

Correlation

Definition:

Let X and Y be two RVs. The *correlation coefficient* is

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

if Var(X)>0 and Var(Y)>0. Otherwise, $\rho(X,Y) = 0$

Note: $-1 \leq \rho(X,Y) \leq 1$

Suppose the covariance of the RVs X and Y is equal to -2.4, and Var(X)=4, Var(Y)=9. Compute $\rho(-6X+2, -3Y)$**.**

A) 0.4

B) - 0.4

C) 7.2

D) -7.2

Exercise

One wants to test the blood of 1000 persons to see which ones are infected by a rare disease. The probability that the test is positive is p=0.001. For efficiency reasons the following procedure is followed:

1. Distribute the blood of the 1000 persons over 25 groups of size 40
2. Mix half of the blood of each of the 40 persons with that of the others in each group
3. Test the aggregated blood sample of each group: when the test is negative, no one in the group is infected. When the test is positive, at least one person in the group is infected, and one will test the other half of the blood of all 40 persons in that group.

Let X_i denote the total numbers of tests needed for the ith group using this procedure.

(a) Describe the probability function of X_i .

(b) Compute E[X_i]

(c) Compute the expected total number of tests

Exercise

Let X and Y be random variables.

(a) Express $\text{Cov}(X, X+Y)$ in terms of $\text{Var}(X)$ and $\text{Cov}(X,Y)$

(b) Are X and $X+Y$ positively correlated, uncorrelated, or negatively correlated?

(c) Same question as in part (b), but now assume that X and Y are uncorrelated.

Sum of 2 independent discrete RVs

Theorem:
Let X and Y be two independent discrete RVs. The p.m.f. of their sum Z=X+Y satisfies

$$p_Z(c) = \sum_j p_X(c - b_j) p_Y(b_j)$$

where the sum runs over all possible values b_j of Y.

Application:
If $X \sim Bin(n, p)$, $Y \sim Bin(n, p)$ and X and Y are independent, then $X + Y \sim Bin(n + m, p)$.

Sum of 2 independent continuous RVs

Theorem:
Let X and Y be two independent continuous RVs. The p.d.f. of their sum Z=X+Y satisfies

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\,\mathrm{d}y$$

Application:

If $X \sim N(\mu, \sigma^2)$, $Y \sim N(\nu, \tau^2)$ and X and Y are independent, then $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$.

Exercise

Let X and Y be independent Poisson distributed random variables with parameters λ and μ.

Show that for $k = 0, 1, 2 \cdots$

$$P(X+Y=k) = \frac{(\lambda+\mu)^k}{k!} e^{-(\lambda+\mu)},$$

by using

$$\sum_{j=0}^{k} \binom{k}{j} p^l (1-p)^{k-j} = 1 \text{ for } p = \mu/(\lambda+\mu)$$

Conclude that the sum of two Poisson RVs is again Poisson, with parameter the sum of the parameters of the original variables.

For next class (week 3.4 lesson 2):

Complete MyStatLab assignments and book exercises

Watch prelectures *'Meet the Poisson process'*

Book: Section 6.1 and 12.1

After this class you

- **know the properties of the Poisson process**
- **know where the Poisson process can serve as model**
- **can simulate random variables from large class of**
- **distributions using standard uniform random variables**

Probability

Good luck!