Probability and random variable theory

Stochastic algorithms

Francesco Corona

Department of Computer Science
Federal University of Ceará, Fortaleza

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws

Probability and random variable theory

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws

We introduce basic probability through the frequency interpretation[1]

- The treatment is non-measure theoretic

Foundation to this interpretation is some kind of **experiment**

- The experiment can be repeatedly performed
- The conditions are seemingly identical

As a result we have a series of **outcomes**

[1]Note that it is here irrelevant whether prior probabilities are assigned from relative frequency, by counting equilikely alternatives or by subjectively quantifying belief. Our focus is to deduce rules for doing operations on probabilities starting from axioms.

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

Each **trial** of the experiment may have any number of specific outcomes

- Say, $1, 2, 3, \cdots, i, \cdots$

Each outcome i either occurs or does not occur (on each trial)

Other legitimate outcomes

- $\underline{i}$, '`NOT` i' (the non-occurrence of i)
- $i \wedge j$, 'i `AND` j' (occurrence of both i and j)
- $i \vee j$, 'i `OR` j' (occurrence of either i or j or both)

Compounding of outcomes can be extended

$$i \wedge (j \wedge k) = (j \wedge i) \wedge k = \cdots = i \wedge j \wedge k$$
$$i \vee (j \vee k) = (j \vee i) \vee k = \cdots = i \vee j \vee k$$

We can also mix $\wedge$ and $\vee$, though those cases are complex

Probability and probability laws (cont.)

Consider n trials of some experiment

Let $m_n(i)$ the number of occurrences of outcome i

We define

$$p(i) = \lim_{n \to \infty} \frac{m_n(i)}{n} \quad (1)$$

If the limit exists, it is called the **probability** of outcome i

- (With respect to that experiment)

'The probability that outcome i will occur'[2]

[2] We can of course replace i with any compound outcome.

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

Suppose that, in the series of n trials, two outcomes i and j occur together
$\rightsquigarrow$ $m_n(i \wedge j)$ times

We have,

$$p(j|i) = \lim_{n \to \infty} \frac{m_n(i \wedge j)}{m_n(i)} \quad (2)$$

If the limit exists, it is called the **conditional probability** of outcome j given ($|$, or conditioned on) outcome i

'The probability that outcome j will occur given that outcome i occurs'

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

$$p(j|i) = \lim_{n \to \infty} \frac{m_n(i \wedge j)}{m_n(i)}$$

If i never occurs, then $p(j|i)$ is undefined
$\rightsquigarrow$ As $m_n(i) = 0$, for all n

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

The set of outcomes (say, $1, 2, 3, \cdots, N$) is said to be **mutually exclusive** if and only if not more than one of them can occur on the same trial

In this case, for any number of trials n

$$\rightsquigarrow \quad m_n(i \wedge j) = 0, \text{ for all } 1 \leq i \leq j \leq N \quad (3)$$

We used the definition $p(i) = \lim_{n \to \infty} \dfrac{m_n(i)}{n}$

Laws of probability (cont.)

A set of outcomes (say, $1, 2, 3, \cdots, N$) is said to be **collectively exhaustive** if and only if at least one of them occurs at every trial

In this case, for any number of trials n

$$\rightsquigarrow \quad p(1 \vee 2 \vee \cdots \vee N) = 1 \tag{4}$$

We used the definition $p(i) = \lim_{n \to \infty} \dfrac{m_n(i)}{n}$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

A set of outcomes maybe both mutually exclusive and collectively exhaustive

- At each trial some, but only those, outcomes occur

(i and $\underline{i}$ is a set of such outcomes)

MAY 02, 2018 _ FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Laws of probability (cont.)

The given definitions allow to deduce the three laws of probability

- $\rightsquigarrow$ **Interval law**
- $\rightsquigarrow$ **Addition law**
- $\rightsquigarrow$ **Multiplication law**

MAY 02, 2018 _ FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

Interval law

The probability $p(i)$ of any outcome i is a real number in $[0, 1]$

$$\rightsquigarrow \quad 0 \leq p(i) \leq 1, \text{ for all } i \tag{5}$$

- $\rightsquigarrow$ $p(i) = 0$ corresponds to the case in which i never occurs
- $\rightsquigarrow$ $p(i) = 1$ corresponds to the case in which i always occurs

Since $0 \leq m_n(i) \leq n$, from $0 \leq m_n(i)/n \leq 1$ and $p(i) = \lim_{n \to \infty} \dfrac{m_n(i)}{n}$

- $\rightsquigarrow$ Outcome i is impossible IFF $m_n(i) = 0$, for all n
- $\rightsquigarrow$ Outcome i is certain IFF $m_n(i) = n$, for all n

■

MAY 02, 2018 _ FC

Probability and probability laws (cont.)

Addition law

Consider N mutually exclusive outcomes $1, 2, \cdots, N$

$$\rightsquigarrow \quad p(1 \vee 2 \vee \cdots \vee N) = p(1) + p(2) + \cdots + p(N) \qquad (6)$$

Mutually exclusive, no more than one of the outcomes can occur in one trial

So, we have

$$m_n(1 \vee 2 \vee \cdots \vee N) = m_n(1) + m_n(2) + \cdots + m_n(N), \text{ for all } n$$

Divide by n and take the limit $n \to \infty$ with the definition of $p(i)$

$\blacksquare$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems
MAY 02, 2018 — FC

Probability and probability laws (cont.)

Mutliplication law

Consider any two outcomes i and j

We have,

$$\rightsquigarrow \quad p(i \wedge j) = p(i)p(j|i) = p(j)p(i|j) \qquad (7)$$

Use the definition $p(j|i) = \lim_{n\to\infty} \dfrac{m_n(i \wedge j)}{m_n(i)}$ and divide both numerator and denominator by n, then take the limit $n \to \infty$

$$p(j|i) = \lim_{n\to\infty} \frac{m_n(i \wedge j)}{m_n(i)} = \lim_{n\to\infty} \frac{m_n(i \wedge j)/n}{m_n(i)/n} = \frac{p(i \wedge j)}{p(i)}$$

$\blacksquare$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems
MAY 02, 2018 — FC

Probability and probability laws (cont.)

Consider the definition of probability and conditional probability

$$p(i) = \lim_{n\to\infty} \frac{m_n(i)}{n}$$
$$p(j|i) = \lim_{n\to\infty} \frac{m_n(i \wedge j)}{m_n(i)}$$

Consider the three laws of probability

$$0 \le p(i) \le 1$$
$$p(1 \vee 2 \vee \cdots \vee N) = p(1) + p(2) + \cdots + p(N)$$
$$p(i \wedge j) = p(i)p(j|i) = p(j)p(i|j)$$

They remain valid if all probabilities therein are conditioned on outcome o

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems
MAY 02, 2018 — FC

Probability and probability laws (cont.)

Events $1, 2, \cdots, N$ are assumed mutually exclusive when o occurs

$$0 \le p(i|o) \le 1$$
$$p(1 \vee 2 \vee \cdots \vee N|o) = p(1|o) + p(2|o) + \cdots + p(N|o)$$

Similarly, also the third equation can be generalised

$$p(i \wedge j|o) = p(i|o)p(j|i \wedge o) = p(j|o)p(i|j \wedge o)$$

Probability and probability laws (cont.)

Consider the definition of probability and conditional probability

$$p(i) = \lim_{n \to \infty} \frac{m_n(i)}{n}$$
$$p(j|i) = \lim_{n \to \infty} \frac{m_n(i \wedge j)}{m_n(i)}$$

Consider the three laws of probability

$$0 \leq p(i) \leq 1$$
$$p(1 \vee 2 \vee \cdots \vee N) = p(1) + p(2) + \cdots + p(N)$$
$$p(i \wedge j) = p(i)p(j|i) = p(j)p(i|j)$$

We can combine the definition of mutual exclusivity $m_n(i \wedge j) = 0$ (for all $1 \leq i \leq j \leq N$) with the multiplication law $p(i \wedge j) = p(i)p(j|i) = p(j)p(i|j)$

- We assume that neither $p(i)$ nor $p(j)$ vanishes

Then, we deduce

$$p(i \wedge j) = 0 \quad \rightsquigarrow p(i|j) = 0 \quad \rightsquigarrow p(j|i) = 0$$
$$p(j|i) = 0 \quad \rightsquigarrow p(i|j) = 0 \quad \rightsquigarrow p(i \wedge j) = 0$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

The addition law deals merely with outcomes that are mutually exclusive

- The multiplicative laws does not

The addition law can still be generalised to non-mutual-exclusiveness

MAY 02, 2018 _ FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

Consider the simple case of $N = 2$ outcomes

Outcome 1 is equivalent to either outcome $1 \wedge 2$ or outcome $1 \wedge \underline{2}$

- They are mutually exclusive

$$p(1) = [p(1 \wedge 2) \vee (1 \wedge \underline{2})] = p(1 \wedge 2) + p(1 \wedge \underline{2})$$

Similarly,

$$p(2) = p(1 \wedge 2) + p(\underline{1} \wedge 2)$$

Outcome $1 \vee 2$ is equivalent to either outcome $1 \wedge 2$ or $1 \wedge \underline{2}$ or $\underline{1} \wedge 2$

- They are mutually exclusive

$$\begin{aligned} p(1 \vee 2) &= p(1 \wedge \underline{2}) + p(1 \wedge 2) + p(\underline{1} \wedge 2) \\ &= p(1) + p(2) - p(1 \wedge 2) \end{aligned} \tag{8}$$

MAY 02, 2018 _ FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Probability and probability laws (cont.)

Consider three outcomes 1, 2 and 3 (not necessarily mutually exclusive)

We list all the implications of the multiplication law

$\rightsquigarrow$ Pairwise relationships, valid for (i, j) any pair of $(1, 2, 3)$

$$p(i \wedge j) = p(i)p(j|i) = p(j)p(i|j)$$

$\rightsquigarrow$ Tertiary relationships, valid for (i, j, k) any permutation of $(1, 2, 3)$

$$p(1 \wedge 2 \wedge 3) = p(i \wedge j)p(k|i \wedge j) \tag{9a}$$
$$p(1 \wedge 2 \wedge 3) = p(i)p(j \wedge k|i) \tag{9b}$$

$\rightsquigarrow$ A combination of the first and second relationships

$$p(1 \wedge 2 \wedge 3) = p(i)p(j|i)p(k|i \wedge j) \tag{10}$$

MAY 02, 2018 _ FC

Probability and probability laws (cont.)

Statistical independence of outcomes (1)

A set of N outcomes is called **statistically independent** if and only if the probability of any one of the outcomes i, conditioned on any other outcome or any 'AND-ed' combination of the other outcomes, is equal to $p(i)$

■

Probability and probability laws (cont.)

For the set of three outcomes 1, 2 and 3, we have

$$\rightsquigarrow \quad p(i) = p(i|j) = p(i|j \wedge k) \qquad (11)$$

This is valid for (i, j, k) all permutations of $(1, 2, 3)$

By using the implications of the multiplication law, we also have

$$\rightsquigarrow \quad p(i \wedge j) = p(i)p(j) \wedge p(1 \wedge 2 \wedge 3) = p(1)p(2)p(3) \qquad (12)$$

This is valid for (i, j) all pairs of $(1, 2, 3)$

'Outcomes 1, 2 and 3 are statistically independent'

Probability and probability laws (cont.)

Statistical independence of outcomes (2)

A set of N outcomes is **statistically independent** if and only if the probability of every 'AND-ed' combination of these outcomes is equal to the product of their separate probabilities $p(i)$

■

Random variables

Probability and random variable theory

Random variables

A **variable** is an entity that always has a value

- We can measure or sample the value

We only consider variables that are real

- Its possible values are real numbers

The value of a variable depends on the context in which it is sampled

Random variables (cont.)

Let variable X be the position at time t of an harmonic oscillator

- Suppose that the intrinsic frequency is ν
- Suppose that the initial position was A
- Suppose that the initial velocity was 0

The value of X will always be found at $A\cos(2\pi\nu t)$

For variable X the sampling context uniquely determines its value

- We call variable X a **sure variable**
- (Your usual variable, that is)

Random variables (cont.)

Suppose that the sampling context does not permit a unique determination

- We (are tempted to) call variable X a **random variable**

The true/correct definition requires more than value uncertainty

General concepts

Random variables

General concepts

We say that X is a (real) random variable if and only if there is a function P of a (real) variable x such that $P(x)\Delta x$ equals (to the first-order in Δx) the probability of finding the value of X in the interval $(x, x + \Delta x]$

That is,

$$\rightsquigarrow \quad \text{Prob}\{X \in (x + \Delta x]\} = P(x)\Delta x + o(\Delta x) \tag{13}$$

$o(\Delta x)$ denotes terms such that $o(\Delta x)/\Delta x \to 0$ as $\Delta x \to 0$

General concepts (cont.)

$$\text{Prob}\{X \in (x + \Delta x]\} = P(x)\Delta x + o(\Delta x)$$

Divide through by Δx, let Δx tend to some infinitesimal positive value $\mathrm{d}x$

$$\rightsquigarrow \quad P(x) = \text{Prob}\{X \in (x, x + \mathrm{d}x]\}/\mathrm{d}x \tag{14}$$

Or, equivalently

$$\rightsquigarrow \quad P(x)\mathrm{d}x = \text{Prob}\{X \in (x, x + \mathrm{d}x]\} \tag{15}$$

'$P(x)$ is the density of probability of the random variable X at value x'

We call P the **density function** of X

General concepts (cont.)

$$\text{Prob}\{X \in (x + \Delta x]\} = P(x)\Delta x + o(\Delta x)$$
$$P(x) = \text{Prob}\{X \in (x, x + \mathrm{d}x]\}/\mathrm{d}x$$
$$P(x)\mathrm{d}x = \text{Prob}\{X \in (x, x + \mathrm{d}x]\}$$

The three equations are equivalent definitions of P

'The RV X is randomly distributed according to the density function P'

General concepts (cont.)

The functional form of P fully defines the random variable X

- The vice versa is also true

Let two RVs X and Y have the same density function P

$$\rightsquigarrow \quad X = Y$$

This does not mean that X and Y have identical values

General concepts (cont.)

Calculate the probability that a sample of X will be in a finite interval $[a, b)$

Imagine the interval to be divided into a set of non-overlapping subintervals

$$\rightsquigarrow \quad [x, x + \mathrm{d}x)$$

$\rightsquigarrow$ The intervals are infinitesimal

The outcome $X \in [a, b)$ can be realised IFF X is inside one such subinterval

- As non-overlapping, X cannot be found in more than one subinterval

By addition law, compute $\text{Prob}\{X \in [a, b)\}$ by summing over subintervals

$$\rightsquigarrow \quad \text{Prob}\{X \in [a, b)\} = \int_a^b P(x)\mathrm{d}x \qquad (16)$$

General concepts (cont.)

Any sample value of X will always lie somewhere in $(-\infty, \infty)$

Thus, we have,

$$\rightsquigarrow \quad \int_{-\infty}^{\infty} P(x)\mathrm{d}x = 1 \qquad (17)$$

We call this equation the **normalisation** or **closure condition**

$\mathrm{d}x$ is intrinsically positive and $P(x)\mathrm{d}x$ is a probability (in $[0, 1]$)

Thus, by the range law

$$\rightsquigarrow \quad P(x) \geq 0, \text{ for all } x \qquad (18)$$

General concepts (cont.)

$$\int_{-\infty}^{\infty} P(x)\mathrm{d}x = 1$$

$$P(x) \geq 0, \text{ for all } x$$

Any such function P can be regarded as a density function

- It will define some random variable X

General concepts (cont.)

We cannot tell for sure what value a RV X will have on a individual sampling

Yet, in the limit of infinitely many samplings, we have

$$P(x)\mathrm{d}x = \text{Prob}\{X \in (x, x + \mathrm{d}x]\}$$

Thus, a fraction $P(x)\mathrm{d}x$ will yield values in $[x, x + \mathrm{d}x)$

A normalised histogram of the sample values of X will approach curve $P(x)$

- (As the number of samples approaches infinity)

General concepts (cont.)

Consider the **Dirac delta function** $\delta(x-x_0)$

Consider the following definition, for any function $f(x)$

$$\int_{-\infty}^{\infty} f(x)\delta(x-x_0)\mathrm{d}x = f(x_0) \tag{19}$$

Consider the following definition, as pair of equations

$$\delta(x-x_0) = 0, \text{ if } x \neq x_0 \tag{20a}$$

$$\int_{-\infty}^{\infty} \delta(x-x_0)\mathrm{d}x = 1 \tag{20b}$$

Compare function $\delta(x-x_0)$ with the definition of a density function

$$\int_{-\infty}^{\infty} P(x)\mathrm{d}x = 1$$

$$P(x) \geq 0, \text{ for all } x$$

$\leadsto$ Function $\delta(x-x_0)$ satisfies all the requirements

$\leadsto$ Function $\delta(x-x_0)$ defines a random variable X

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

General concepts (cont.)

Take into account the closure condition, we have for random variable X

$$\text{Prob}\{X \in [a,b)\} = \int_a^b \underbrace{\delta(x-x_0)}_{P(x)}\mathrm{d}x = \begin{cases} 1, & x_0 \in (a,b] \\ 0, & x_0 \notin (a,b] \end{cases}$$

The probability of finding the value of X inside any interval is either 0 or 1

- This depends on whether that interval does or does not contain x_0

$\leadsto$ Thus, the value of X must be precisely x

'*Random variable X with density function $\delta(x-x_0)$ is the sure variable x_0*'

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

General concepts (cont.)

The reverse implication

'*The sure variable x_0 has density function $\delta(x-x_0)$*

The Dirac delta function $\delta(x-x_0)$ is the only function of x whose integral from $x=a$ to $x=b$ is unity if $x_0 \in [a,b)$ and zero otherwise

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

General concepts (cont.)

The **distribution function** is also a function of a random variable X

- It is closely related to the density function of RV X

The definition

$$F(x) \equiv \text{Prob}\{X < x\} \tag{21}$$

By combining the definition with $\text{Prob}\{X \in [a,b)\} = \int_a^b P(x)\mathrm{d}x$, we get

$$\leadsto \quad F(x) = \int_{-\infty}^{x} P(x')\mathrm{d}x' \tag{22}$$

By differentiation with respect to x, we get

$$\leadsto \quad P(x) = F'(x) \tag{23}$$

General concepts (cont.)

The integral-derivative relationship between function F and function P

The probability that a sample value of X will be in $[x, x + \mathrm{d}x)$

- $\rightsquigarrow$ $\text{Prob}\{X \in [x, x + \mathrm{d}x)\} = P(x)\mathrm{d}x$
- $\rightsquigarrow$ $\mathrm{d}F(x) = F'(x)\mathrm{d}x = P(x)\mathrm{d}x$

The probability that a sample value of X will be less than x

General concepts (cont.)

$$F(x) = \int_{-\infty}^{x} P(x')\mathrm{d}x'$$

$$\int_{-\infty}^{\infty} P(x)\mathrm{d}x = 1$$

$$P(x) \geq 0, \text{ for all } x$$

Function $F(X)$ increases monotonically, from 0 at $x = -\infty$ to 1 at $x = \infty$

Any function with this property can be regarded as distribution function F

- $\rightsquigarrow$ Function F defines a random variable X

General concepts (cont.)

$$P(x) = F'(x)$$

If we know P, we can calculate F and vice versa

- $\rightsquigarrow$ Either one defines the random variable

Moments

Random variables

Moments

Let h be any univariate function

We define the **average** of h

⇝ With respect to the RV X

$$\langle h(X) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} h\big[x^{(i)}\big] \quad (24)$$

$x^{(1)}, x^{(2)}, \dots, x^{(N)}$ are values assumed by X in N independent samplings

Consider the definitions

$$P(x)\mathrm{d}x = \text{Prob}\{X \in (x, x + \mathrm{d}x]\}$$

$$p(i) = \lim_{N\to\infty} \frac{m_N(i)}{N}$$

$P(x)\mathrm{d}x$ is the approximate fraction of any N samples of X in $[x, x + \mathrm{d}x)$

⇝ (The approximation becomes exact as $N \to \infty$)

$NP(x)\mathrm{d}x \times h(x)$ is the contribution to the sum from X values in $[x, x + \mathrm{d}x)$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

Moments (cont.)

$$\langle h(X) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} h\big[x^{(i)}\big]$$

We can approximate the sum

$$\rightsquigarrow \quad \sum_{i=1}^{N} h(x^{(i)}) \approx \int_{x=-\infty}^{\infty} \big[NP(x)\mathrm{d}x \times h(x)\big] = N \int_{-\infty}^{\infty} P(x)\mathrm{d}x \times h(x)$$

Divide by N and take the limit $N \to \infty$ (so to get an equality)

$$\rightsquigarrow \quad \langle h(X) \rangle = \int_{-\infty}^{\infty} h(x)P(x)\mathrm{d}x \quad (25)$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

Moments (cont.)

$$\langle h(X) \rangle = \int_{-\infty}^{\infty} h(x)P(x)\mathrm{d}x$$

The formula gives $\langle h(X) \rangle$ in terms of function h and density P

- This form is analytically more convenient

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

Moments (cont.)

Some useful averages

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n P(x)\mathrm{d}x, \quad (n = 1, 2, \dots) \quad (26)$$

$\langle X^n \rangle$ is called the **n-th moment of X**

Remembering the closure condition, we have

$$\rightsquigarrow \quad \langle X^0 \rangle = \int_{-\infty}^{\infty} P(x)\mathrm{d}x = 1 \quad (27)$$

Higher order moments may or may not exist (usually some exist, at least)

Moments (cont.)

Suppose that all moments of X exist, they determine the average of any h

- (Any function h, as in analytical function, infinitely differentiable)

$$\rightsquigarrow \quad \langle h(X) \rangle = \int_{-\infty}^{\infty} \underbrace{\Big[\sum_{n=0}^{\infty} \frac{h^{(n)}(0)}{n!} x^n \Big]}_{h(x)} P(x) \mathrm{d}x = \sum_{n=0}^{\infty} \frac{h^{(n)}(0)}{n!} \langle X^n \rangle$$

$h^{(n)}(0)$ is the n-th derivative of function h evaluated at $x = 0$

$\rightsquigarrow$ (We used Taylor's series to represent h)

Moments (cont.)

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n P(x) \mathrm{d}x, \quad (n = 1, 2, \dots)$$

Consider a RV X that is not identically zero

- Its $P(x)$ is other than $\delta(x - x_0)$

The integral is strictly positive when n is even

We thus have,

$$X \not\equiv 0 \quad \rightsquigarrow \langle X^{2n} \rangle > 0, \quad (n = 1, 2, \dots) \tag{28}$$

Moments (cont.)

Mean of X

The **mean** of X is defined to be the first ($n = 1$) moment of X

$$\text{mean}\{X\} = \langle X \rangle = \int_{-\infty}^{\infty} x P(x) \mathrm{d}x \tag{29}$$

■

Moments (cont.)

Variance of X

The **variance** of X

$$\text{var}\{X\} = \langle (X - \langle X \rangle)^2 \rangle = \int_{-\infty}^{\infty} (x - \langle X \rangle)^2 P(x) \mathrm{d}x \tag{30}$$

The integral on the RHS

$$\begin{aligned} \int_{-\infty}^{\infty} (x^2 - 2\langle X \rangle x + \langle X^2 \rangle) P(x) \mathrm{d}x \\ = \langle X^2 \rangle - 2 \langle X \rangle \langle X \rangle + \langle X \rangle^2 \\ = \langle X^2 \rangle - \langle X \rangle^2 \end{aligned}$$

The variance of X in terms of the first and the second order moments of X

$$\rightsquigarrow \quad \text{var}\{X\} = \langle X^2 \rangle - \langle X \rangle^2 \tag{31}$$

■

Moments (cont.)

$$\text{var}\{X\} = \langle X^2 \rangle - \langle X \rangle^2$$

Note the non-negativity of the integrand in $\text{var}\{X\} = \int_{-\infty}^{\infty} (x - \langle X \rangle)^2 P(x) \mathrm{d}x$

We have that $\text{var}\{X\} \geq 0$

$$\rightsquigarrow \quad \langle X^2 \rangle \geq \langle X \rangle^2 \tag{32}$$

The equality holds if and only if $\text{var}\{X\} = \langle (X - \langle X \rangle)^2 \rangle = 0$

- It happens if and only if X is the sure variable $\langle X \rangle$
- (Sure variables have zero variance)

'*If* $\text{var}\{X\} = 0$*, then* X *is the sure variable* $X = \langle X \rangle$'

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Moments (cont.)

$$\langle X^2 \rangle \geq \langle X \rangle^2$$

We can observe that $(X - \langle X \rangle)^2$ is strictly positive everywhere

- The only exception being $x = \langle X \rangle$

Moreover, $P(x)$ is never negative

Thus, we find the only possibility for the integral be equal zero

- $P(x)$ must be zero everywhere except possibly at $x = \langle X \rangle$

Because of normalisation, we must have $P(x) = \delta(x - \langle X \rangle)$

- This implies that X is the sure variable $\langle X \rangle$

MAY 02, 2018 — FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Moments (cont.)

The reverse implication

'*The sure variable* $X = \langle X \rangle$ *has* $\text{var}\{X\} = 0$'

We can observe that for a sure variable we have $P(x) = \delta(x - \langle X \rangle)$

By substitution into $\text{var}\{X\} = \int_{-\infty}^{\infty} \underbrace{(x - \langle X \rangle)^2}_{f(x)} \underbrace{\delta(x - \langle X \rangle)}_{P(x)} \mathrm{d}x$

$$\rightsquigarrow \quad \text{var}\{X\} = 0$$

We used $\int_{-\infty}^{\infty} f(x)\delta(x - x_0)\mathrm{d}x = f(x_0)$

MAY 02, 2018 — FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Moments (cont.)

Standard deviation of X

The square root for the variance of X is called the **standard deviation**

The definition

$$\begin{aligned} \text{sdev}\{X\} \equiv (\text{var}\{X\})^{1/2} &\equiv \langle (X - \langle X \rangle)^2 \rangle^{1/2} \\ &= \left[\langle X^2 \rangle - \langle X \rangle^2 \right]^{1/2} \end{aligned} \tag{33}$$

$\text{sdev}\{X\}$ is the square root of the average squared difference between X and its mean $\langle X \rangle$, thus it measures the size of expected difference between the sample values of X and the mean of X

- The size of the expected 'dispersion of' or 'fluctuation in' X about $\langle X \rangle$

MAY 02, 2018 — FC

Moments (cont.)

Complete knowledge of RV X requires knowledge of its density function P

Partial knowledge provided by $\langle X \rangle$ and $\text{sdev}\{X\}$ often suffices, in practice

- $\rightsquigarrow$ $\langle X \rangle$ is the best possible 'sure number approximation' of RV X
- $\rightsquigarrow$ $\text{sdev}\{X\}$ measures how (in-)accurate that approximation is

(The characterisation is neither unique nor complete, though)

Named random variables

Random variables

Named random variables

- $\rightsquigarrow$ The **exponential random variable**
- $\rightsquigarrow$ The **uniform random variable**
- $\rightsquigarrow$ The **normal random variable**
- $\rightsquigarrow$ ...

The uniform RV

The uniform random variable

The **uniform random variable** is defined by the density function $P(x)$

$$P(x) = \begin{cases} 1/(b-a), & a \leq x < b \\ 0, & \text{elsewhere} \end{cases} \quad (34)$$

a and b are two real numbers such that $-\infty < a < b < \infty$

Function P satisfies the closure and the non-negativity condition

$$\int_{-\infty}^{\infty} P(x) \mathrm{d}x = 1$$

$$P(x) \geq 0, \text{ for all } x$$

'*The RV X defined by density function P is uniformly distributed in $[a, b)$*'

$$X \sim \mathcal{U}(a, b)$$

The uniform RV (cont.)

$$P(x) = \begin{cases} 1/(b-a), & a \leq x < b \\ 0, & \text{elsewhere} \end{cases}$$

We can substitute the density function into the moment definition

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n P(x) \mathrm{d}x, \quad (n = 1, 2, \dots)$$

For $X \sim \mathcal{U}(a, b)$, we obtain[3]

$$\rightsquigarrow \quad \langle X^n \rangle = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)} = \frac{1}{(n+1)} \sum_{j=0}^{n} a^j b^{n-j} \tag{35}$$

[3] We used the integral identity, for $n = 0, 1, \dots$

$$\int_a^b \mathrm{d}x x^n = \frac{b^{n+1} - a^{n+1}}{(n+1)} = \frac{b-a}{(n+1)} \sum_{j=0}^{n} a^j b^{n-j}.$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

The uniform RV (cont.)

$$\langle X^n \rangle = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)} = \frac{1}{(n+1)} \sum_{j=0}^{n} a^j b^{n-j}$$

For $X \sim \mathcal{U}(a, b)$, let $n = 1$ and $n = 2$, we get

$$\rightsquigarrow \quad \mathrm{mean}\{X\} = \frac{a+b}{2} \tag{36a}$$

$$\rightsquigarrow \quad \mathrm{sdev}\{X\} = \frac{b-a}{2\sqrt{3}} \tag{36b}$$

We used

$$\mathrm{mean}\{X\} = \langle X \rangle$$
$$\mathrm{sdev}\{X\} = (\langle X^2 \rangle - \langle X \rangle^2)^{1/2}$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

The uniform RV (cont.)

$$\mathrm{mean}\{X\} = \frac{a+b}{2}$$
$$\mathrm{sdev}\{X\} = \frac{b-a}{2\sqrt{3}}$$

Let parameter b approach a from above, we get

$$\mathrm{mean}\{X\} \rightarrow a$$
$$\mathrm{sdev}\{X\} \rightarrow 0$$

We conclude,

$$\lim_{b \rightarrow a} \mathcal{U}(a, b) = a \tag{37}$$

■

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

The exponential RV

The exponential random variable

The **exponential random variable** is defined by the density function $P(x)$

$$P(x) = \begin{cases} a \exp -ax, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases} \tag{38}$$

a is a real numbers such that $a > 0$

Function P satisfies the closure and the non-negativity condition

$$\int_{-\infty}^{\infty} P(x) \mathrm{d}x = 1$$
$$P(x) \geq 0, \text{ for all } x$$

'RV X defined by density function P is exponentially distributed, decay a'

$$X \sim \mathcal{E}(a)$$

The exponential RV (cont.)

$$P(x) = \begin{cases} a \exp - ax, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

We can substitute the density function into the moment definition

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n P(x) \mathrm{d}x, \quad (n = 1, 2, \dots)$$

For $X \sim \mathcal{E}(a)$, we obtain[4]

$$\rightsquigarrow \quad \langle X^n \rangle = \frac{n!}{a^n} \tag{39}$$

[4] We used the integral identity, for $a > 0$ and for $n = 0, 1, \dots$

$$\int_0^{\infty} \mathrm{d}x x^n \exp{(-ax)} = \frac{n!}{a^{n+1}}.$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

The exponential RV (cont.)

$$\langle X^n \rangle = \frac{n!}{a^n}$$

For $X \sim \mathcal{E}(a)$, let $n = 0$, we get the closure condition

$$\langle X^0 \rangle = 1$$

For $X \sim \mathcal{E}(a)$, let $n = 1$ and $n = 2$, we get

$$\rightsquigarrow \quad \text{mean}\{X\} = 1/a \tag{40a}$$

$$\rightsquigarrow \quad \text{sdev}\{X\} = 1/a \tag{40b}$$

We used

$$\text{mean}\{X\} = \langle X \rangle$$

$$\text{sdev}\{X\} = (\langle X^2 \rangle - \langle X \rangle^2)^{1/2}$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

The exponential RV (cont.)

$$\text{mean}\{X\} = 1/a$$

$$\text{sdev}\{X\} = 1/a$$

Let parameter a approach infinity, we get

$$\text{mean}\{X\} \to 0$$

$$\text{sdev}\{X\} \to 0$$

We conclude,

$$\lim_{a \to \infty} \mathcal{E}(a) = 0 \tag{41}$$

■

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

The normal RV

The normal random variable

The **normal random variable** is defined by the density function $P(x)$

$$P(x) = \frac{1}{(2\pi a^2)^{1/2}} \exp\left[-\frac{(x-m)^2}{2a^2}\right] \tag{42}$$

a and m are real numbers such that $0 < a < \infty$ and $-\infty < m < \infty$

Function P satisfies the closure and the non-negativity condition

$$\int_{-\infty}^{\infty} P(x) \mathrm{d}x = 1$$

$$P(x) \geq 0, \quad \text{for all } x$$

'RV X defined by density function P is normally distributed, mean m and variance a^2'

$$X \sim \mathcal{N}(m, a^2)$$

The normal RV (cont.)

$$P(x) = \frac{1}{(2\pi a^2)^{1/2}} \exp\Big[-\frac{(x-m)^2}{2a^2}\Big]$$

We can substitute the density function into the moment definition

$$\langle X^n \rangle = \int_{-\infty}^{\infty} x^n P(x) \mathrm{d}x, \quad (n = 1, 2, \dots)$$

For $X \sim \mathcal{N}(m, a^2)$, we obtain[5]

$$\rightsquigarrow \quad \langle X^n \rangle = n! \sum_{\substack{k=0 \\ (\text{even})}}^{n} \frac{m^{n-k}(a^2)^{k/2}}{(n-k)!(k/2)!2^{k/2}} \qquad (43)$$

[5] We changed the integration variable from x to $z = (x-m)/a\sqrt{2}$, and we used the binomial formula

$$(x+y)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^{n-k} y^k.$$

We also used the integral identity

$$\int_{-\infty}^{\infty} \mathrm{d}x x^n \exp(-a^2x^2) = \begin{cases} \dfrac{n!}{(n/2)!(2a)^n} \dfrac{\pi^{1/2}}{|a|}, & n = 0, 2, 4, \dots \\ 0, & n = 1, 3, 5, \dots \end{cases}.$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

The normal RV (cont.)

$$\langle X^n \rangle = n! \sum_{\substack{k=0 \\ (\text{even})}}^{n} \frac{m^{n-k}(a^2)^{k/2}}{(n-k)!(k/2)!2^{k/2}}$$

For $X \sim \mathcal{N}(m, a^2)$, let $n = 0$, we get the closure condition

$$\langle X^0 \rangle = 1$$

For $X \sim \mathcal{N}(m, a^2)$, let $n = 1$ and $n = 2$, we get

$$\rightsquigarrow \quad \text{mean}\{X\} = m \qquad (44a)$$
$$\rightsquigarrow \quad \text{sdev}\{X\} = a \qquad (44b)$$

We used

$$\text{mean}\{X\} = \langle X \rangle$$
$$\text{sdev}\{X\} = (\langle X^2 \rangle - \langle X \rangle^2)^{1/2}$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

The normal RV (cont.)

$$\text{mean}\{X\} = m$$
$$\text{sdev}\{X\} = a$$

Let parameter a approach to zero, we get

$$\text{mean}\{X\} \to m$$
$$\text{sdev}\{X\} \to 0$$

We conclude,

$$\lim_{a \to 0} \mathcal{N}(m, a^2) = m \qquad (45)$$

■

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1

Discrete variables

Random variables

Discrete variables

It is convenient to define RVs that can only take on discrete/integer values

A RV X is said to be a **discrete random variable** if and only if there exists a function P of an integer value n such that $P(n)$ equals the probability of finding the value of X to be n

$$P(n) \equiv \text{Prob}\{X = n\} \tag{46}$$

We again call P the **density function** of X

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

Calculate the probability of finding a value of X between n_1 and n_2, $[n_1, n_2]$

The outcome can be realised IFF X is one such integer number

By the addition law of probability,

$$\rightsquigarrow \quad \text{Prob}\{X \in [n_1, n_2]\} = \sum_{n=n_1}^{n_2} P(n) \tag{47}$$

We summed over the integers between n_1 and n_2

MAY 02, 2018 — FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

Any sample value of X will surely take on some integer value in $(-\infty, \infty)$

Thus, we have,

$$\rightsquigarrow \quad \sum_{n=-\infty}^{\infty} P(n) = 1 \tag{48}$$

We call this equation the **normalisation** or **closure condition**

$P(n)$ is a probability (in $[0, 1]$), thus by the range law

$$\rightsquigarrow \quad 0 \geq P(n) \geq 1, \quad \text{for all } n \tag{49}$$

MAY 02, 2018 — FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

$$\sum_{n=-\infty}^{\infty} P(n) = 1$$

$$P(n) \in [0, 1], \quad \text{for all } n$$

Any such function P can be regarded as a density function

- It will define some discrete random variable X

MAY 02, 2018 — FC

Discrete variables (cont.)

Consider the **Kronecker delta function** $\delta(n, n_0)$

Suppose that $\delta(n, n_0)$ is a density function $P(n)$

$$P(n) = \delta(n, n_0) = \begin{cases} 1, & n = n_0 \\ 0, & n \neq n_0 \end{cases} \tag{50}$$

From the definition of a density function, we have

$$\sum_{n=-\infty}^{\infty} P(n) = 1$$
$$P(n) \in [0, 1], \text{ for all } n$$

We have that function $\delta(n, n_0)$ satisfies all the requirements

$\rightsquigarrow$ Function $\delta(n, n_0)$ defines a discrete random variable X

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

Take into account the closure condition, we have for the discrete RV X

$$\rightsquigarrow \quad \text{Prob}\{X = n_0\} = \sum_{n=n_0}^{n_0} \delta(n, n_0) = \begin{cases} 1, & n = n_0 \text{ (always)} \\ 0, & n \neq n_0 \text{ (never)} \end{cases}$$

The probability of finding the value of X to be n_0 is either 0 or 1

- To be one, the value of n must be n_0

'*Random variable X with density function $\delta(n, n_0)$ is the sure variable n_0*'

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

We define the **distribution function** F of a discrete random variable X

$$F(n) \equiv \text{Prob}\{X \leq n\} \tag{51}$$

Combining the definition with $\text{Prob}\{X \in [n_1, n_2]\} = \sum_{n=n_1}^{n_2} P(n)$, we get

$$\rightsquigarrow \quad F(n) = \sum_{n'=-\infty}^{n} P(n') \tag{52}$$

This implies

$$\rightsquigarrow \quad P(n) = F(n) - F(n-1) \tag{53}$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

$$F(n) = \sum_{n'=-\infty}^{\infty} P(n')$$

Function $F(x)$ increases monotonically from 0 at $n = -\infty$ to 1 at $n = \infty$

If we know $P(x)$, we can calculate $F(x)$ and vice versa

- Either one define the discrete random variable

$$\sum_{n=-\infty}^{\infty} P(n) = 1$$
$$P(n) \in [0, 1], \text{ for all } n$$

Discrete variables (cont.)

Probability and random variable theory

UFC/DC
SA (CK0191)
2018.1

Integer numbers are a subset of the real numbers

- Are discrete RVs a special case of real RVs?

Discrete variables (cont.)

Probability and random variable theory

UFC/DC
SA (CK0191)
2018.1

Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Consider as real random variable Y with given density function $Q(Y)$

$$Q(y) = \sum_{n=-\infty}^{\infty} P(n)\delta(y-n) \quad (54)$$

⇝ $P(n)$ is the density function of the integer random variable X

$$P(n) \equiv \text{Prob}\{X = n\}$$

⇝ $\delta(y-n)$ is the Dirac delta function of real variable y

From $\text{Prob}\{Y \in [a,b)\} = \int_a^b P(x)\text{d}x$, we obtain

$$\rightsquigarrow \quad \text{Prob}\{Y \in [a,b)\} = \int_a^b Q(y)\text{d}y = \sum_{-\infty}^{\infty} P(n) \int_a^b \delta(y-n)\text{d}y$$

Discrete variables (cont.)

Probability and random variable theory

UFC/DC
SA (CK0191)
2018.1

Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

$$\text{Prob}\{Y \in [a,b)\} = \int_a^b Q(y)\text{d}y = \sum_{-\infty}^{\infty} P(n) \int_a^b \delta(y-n)\text{d}y$$

- The probability of finding Y inside any interval that does not contain at least one integer is zero
- The probability of finding Y inside any interval that contains one integer n' is $P(n')$

This implies that the real random variable Y with density Q is identical to the discrete random variable X with density P

Discrete variables (cont.)

Probability and random variable theory

UFC/DC
SA (CK0191)
2018.1

Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

The theory of discrete RVs is not very different from the theory of real RVs

⇝ Basic relations are obtained by replacing integration with summation

⇝ (And, Dirac delta function with Kronecker delta functions)

Discrete variables (cont.)

Consider the definition of **average** of function h with respect to real RV Y

$$\langle h(Y) \rangle = \int_{-\infty}^{\infty} h(y) Q(y) \mathrm{d}x$$

We want to adapt the definition of average to the discrete RV case

Let Y be the real RV with density function Q

$$Q(y) = \sum_{n=-\infty}^{\infty} P(n) \delta(y-n)$$

The integer RV X with density P is identical

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

For the average of h with respect to RV Y, we have

$$\begin{aligned}\langle h(Y) \rangle &= \int_{-\infty}^{\infty} h(y) Q(y) \mathrm{d}y \\ &= \sum_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y) P(n) \delta(y-n) \mathrm{d}y \\ &= \sum_{n=-\infty}^{\infty} h(n) P(n) \rightsquigarrow \langle h(X) \rangle\end{aligned}$$

(As random variables Y and X are identical)

$$\rightsquigarrow \quad \langle h(X) \rangle = \sum_{n=-\infty}^{\infty} h(n) P(n) \tag{55}$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

We define the **k-th moment** of the discrete RV X

$$\langle X^k \rangle = \sum_{n=-\infty}^{\infty} n^k P(n) \tag{56}$$

We can use this expression to calculate mean and variance of X

- Again, the vanishing $\text{var}\{X\}$ is a necessary and sufficient condition for X to be the sure variable $\langle X \rangle$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

The discrete uniform variable

The **discrete uniform variable** is defined by the density function $P(n)$

$$P(n) = \begin{cases} (n_2 - n_1 + 1)^{-1}, & n_1 \leq n \leq n_2 \\ 0, & n < n_1 \text{ or } n > n_2 \end{cases} \tag{57}$$

n_1 and n_2 are any two integers such that $n_1 \leq n_2$

For this RV integer values between n_1 and n_2 inclusively are equally probable

Discrete variables (cont.)

$$P(n) = \begin{cases} (n_2 - n_1 + 1)^{-1}, & n_1 \leq n \leq n_2 \\ 0, & n < n_1 \text{ or } n > n_2 \end{cases}$$

We can substitute the density function P into the moment definition

$$\langle X^k \rangle = \sum_{n=-\infty}^{\infty} n^k P(n)$$

The results[6], for $k=1$ and $k=2$

$$\rightsquigarrow \quad \langle X \rangle = (n_1 + n_2)/2 \tag{58a}$$

$$\rightsquigarrow \quad \mathrm{var}\{X\} = (n_2 - n_1)(n_2 - n_1 + 2)/12 \tag{58b}$$

For $n_2 = n_1$, we have $\langle X \rangle = n_1$ and $\mathrm{var}\{X\} = 0$ (X is the sure variable n_1)

■

[6] We used two algebraic identities,

$$\sum_{n=1}^{N} n = N(N+1)/2$$
$$\sum_{n=1}^{N} n^2 = N(N+1)(2N+1)/6 \quad .$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

The binomial variable

The **binomial variable** is defined by the density function $P(n)$

$$P(n) = \begin{cases} \dfrac{N!}{n!(N-n)!} p^n (1-p)^{N-n}, & 0 \geq n \geq N \\ 0, & n < 0 \text{ or } n > N \end{cases} \tag{59}$$

N is any positive number and p is any real number such that $0 \leq p \leq 1$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

$$P(n) = \begin{cases} \dfrac{N!}{n!(N-n)!} p^n (1-p)^{N-n}, & 0 \geq n \geq N \\ 0, & n < 0 \text{ or } n > N \end{cases}$$

We can substitute the density function into the moment definition

$$\langle X^k \rangle = \sum_{n=-\infty}^{\infty} n^k P(n)$$

The results[7] for $k=1$ and $k=2$

$$\rightsquigarrow \quad \langle X \rangle = Np \tag{60a}$$

$$\rightsquigarrow \quad \mathrm{var}\{X\} = Np(1-p) \tag{60b}$$

- $\rightsquigarrow$ For $p=0$, X is the sure variable 0
- $\rightsquigarrow$ For $p=1$, X is the sure variable N

■

[7] We used the algebraic identity,

$$1 = [p + (1-p)]^n = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}.$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

The Poisson variable

The **Poisson variable** is defined by the density function $P(n)$

$$P(n) = \begin{cases} \dfrac{e^{-a} a^n}{n!}, & n \geq 0 \\ 0, & n < 0 \end{cases} \tag{61}$$

a is any positive real number

Discrete variables (cont.)

$$P(n) = \begin{cases} \dfrac{e^{-a}a^n}{n!}, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

We can substitute the density function into the moment definition

$$\langle X^k \rangle = \sum_{n=-\infty}^{\infty} n^k P(n)$$

The results[8] for $k = 1$ and $k = 2$

$$\rightsquigarrow \quad \langle X \rangle = a \qquad (62a)$$

$$\rightsquigarrow \quad \text{var}\{X\} = a \qquad (62b)$$

For $p \to 0$, the Poisson RV approaches the sure variable 0

[8] We used the algebraic identity,

$$e^a = \sum_{n=0}^{\infty} \frac{a^n}{n!}.$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

The Poisson RV can be understood as a special case of the binomial RV

- For N very large and p very small (Rare events)

$$P(n) = \begin{cases} \dfrac{e^{-a}a^n}{n!}, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

Let N be large and approximate the density function of the binomial RV

$$P(n) = \begin{cases} \dfrac{N!}{n!(N-n)!} p^n (1-p)^{N-n}, & 0 \geq n \geq N \\ 0, & n < 0 \text{ or } n > N \end{cases}$$

We have,

$$\frac{N!}{(N-n)!} p^n = N(N-1)\cdots(N-n+1)p^n \approx N^n p^n = (Np)^n$$

$$(1-p)^{N-n} \approx (1-p)^N = \Big(1 - \frac{Np}{N}\Big)^N \approx e^{-Np}$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

$$\frac{N!}{(N-n)!} p^n = N(N-1)\cdots(N-n+1)p^n \approx N^n p^n = (Np)^n$$

$$(1-p)^{N-n} \approx (1-p)^N = \Big(1 - \frac{Np}{N}\Big)^N \approx e^{-Np}$$

As a result,

$$\frac{N!}{(N-n)!} p^n (1-p)^{N-n} \quad \rightsquigarrow \overset{N \to \infty}{p = a/N} \rightsquigarrow \quad \frac{e^{-a}a^n}{n!} \qquad (63)$$

For $N \to \infty$ and $p \to 0$ with Np constant, the binomial RV with parameters N and p becomes the Poisson RV with mean and variance Np

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

Consider the Poisson RV with density function $P(n)$

$$P(n) = \begin{cases} \dfrac{e^{-a}a^n}{n!}, & n \geq 0 \\ 0, & n < 0 \end{cases}$$

For $a >> 1$, $P(n)$ is appreciably larger than zero only for $n \approx a$

Consider the Stirling's approximation of $n!$ (?)

$$n! \approx (2\pi n)^{1/2} n^n e^{-n}, \quad \text{for } n >> 1$$

Moreover, we establish

$$\ln\Big[\frac{e^{-a}a^n}{n!}(2\pi a)^{1/2}\Big] \approx (n-a) - n\ln(n/a), \quad \text{for } n >> 1$$

Discrete variables (cont.)

The logarithm approximation

$$\ln(1+\varepsilon) \approx \varepsilon - \varepsilon^2/2, \text{ for } |\varepsilon| << 1$$

$\varepsilon = (n-a)/a$, so when $n \approx a$, the RHS is approximately equal to

$$-a\varepsilon^2/2 = -(n-a)^2/(2a)$$

We conclude

$$\frac{e^{-a}a^n}{n!} \approx \frac{1}{(2\pi a)^{1/2}} \exp\Big[-\frac{(n-a)^2}{2a}\Big], \quad (a >> 1) \qquad (64)$$

When a is sufficiently large, the Poisson RV with mean and variance a approaches (at integer values) the normal RV with mean and variance a

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Discrete variables (cont.)

We may further elaborate

$$\frac{N!}{n!(N-n)!}p^n(a-p)^{N-n} \overset{N\to\infty}{\rightsquigarrow} p = a/N, \ a >> 1$$

$$\rightsquigarrow \quad \frac{1}{(2\pi a)^{1/2}} \exp\Big[-\frac{(n-a)^2}{2a}\Big] \qquad (65)$$

If $N \to \infty$ and $p \to 0$ with Np a sufficiently large constant, then the binomial RV with parameters N and p approaches (at integer values) the normal RV with mean and variance Np

■

MAY 02, 2018 — FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables

Random variables

MAY 02, 2018 — FC

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables

Consider a set of n variables $X_1, X_2, \dots, X_n$

The set is called a set of **joint random variables** if and only if

1. all n variables can be sampled simultaneously
2. there is a n-variate function P such that $P(x_1, x_2, \dots, x_n)\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$ equals the probability that such a sampling finds X_i inside $[x_i, x_i + \mathrm{d}x_i)$ for all $i = 1, 2, \cdots, n$

MAY 02, 2018 — FC

Joint random variables (cont.)

Consider for simplicity the case $n = 3$ random variables X_1, X_2 and X_3

For any sampling, we have

$$P(x_1, x_2, x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = \text{Prob}\{X_i \in [x_i, x_i + \mathrm{d}x_i), \text{ for } i = 1, 2, 3\} \quad (66)$$

Again, the equation is considered valid only to the first-order

- In each of the positive infinitesimals $\mathrm{d}x_1$, $\mathrm{d}x_2$ and $\mathrm{d}x_3$

Function P is called **joint density function** of X_1 ,X_2 and X_3

$\rightsquigarrow$ The expression above is regarded as its definition

'X_1, X_2 and X_3 are distributed according to the joint density function P'

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Individually sampling any one variable X_1, X_2 or X_3 is equivalent to sampling them all simultaneously and then ignoring the values of the other two

- This is implicit in the main working hypothesis
- (X_1, X_2 or X_3 can be sampled simultaneously)

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Because of the addition law of probability, we have

$$\rightsquigarrow \int_{a_1}^{b_1} \mathrm{d}x_1 \int_{a_2}^{b_2} \mathrm{d}x_2 \int_{a_3}^{b_3} \mathrm{d}x_3\, P(x_1, x_2, x_3) = \text{Prob}\{X_i \in [a_i, b_i), \text{ for } i = 1, 2, 3\} \quad (67)$$

Moreover, as each RV X_i is certain to be found in $(-\infty, \infty)$

$$\rightsquigarrow \int_{-\infty}^{\infty} \mathrm{d}x_1 \int_{-\infty}^{\infty} \mathrm{d}x_2 \int_{\infty}^{\infty} \mathrm{d}x_3\, P(x_1, x_2, x_3) = 1 \quad (68)$$

The expression is again called the **normalisation** or **closure condition**

The infinitesimals $\mathrm{d}x_i$ are intrinsically positive

$$\rightsquigarrow \quad P(x_1, x_2, x_3) \geq 0 \quad (69)$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

$$\int_{-\infty}^{\infty} \mathrm{d}x_1 \int_{-\infty}^{\infty} \mathrm{d}x_2 \int_{\infty}^{\infty} \mathrm{d}x_3\, P(x_1, x_2, x_3) = 1$$

$$P(x_1, x_2, x_3) \geq 0$$

Any such three-variable function P can be regarded as joint density function

- It defines the joint random variables X_1, X_2 and X_3

Joint random variables (cont.)

Consider the analogy with the distribution function $F(x) \equiv \text{Prob}\{X < x\}$

We can define the **joint distribution function** of RVs X_1, X_2 and X_3

The definition

$$F(x_1, x_2, x_3) \equiv \text{Prob}\{X_i < x_i, \text{ for } i = 1, 2, 3\} \quad (70a)$$

$$= \int_{-\infty}^{x_1} \mathrm{d}x_1' \int_{-\infty}^{x_2} \mathrm{d}x_2' \int_{-\infty}^{x_3'} \mathrm{d}x_3' P(x_1', x_2', x_3') \quad (70b)$$

We used $\int_{a_1}^{b_1} \mathrm{d}x_1 \int_{a_2}^{b_2} \mathrm{d}x_2 \int_{a_3}^{b_3} \mathrm{d}x_3 P(x_1, x_2, x_3) = \text{Prob}\{X_i \in [a_i, b_i), \forall i\}$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

The joint density function P produces a number of density functions

- They are important

These density functions can be grouped into two main classes

⇝ **Marginal density functions**

⇝ **Conditional density functions**

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Let (i, j, k) denote any permutation of $(1, 2, 3)$

We define the **marginal density functions**

$$P_i(x_i)\mathrm{d}x_i \equiv \text{Prob}\{X_i \in [x_i, x_i + \mathrm{d}x_i), \text{ regardless of } X_j \text{ and } X_k\} \quad (71)$$

$$P_{i,j}(x_i, x_j)\mathrm{d}x_i\mathrm{d}x_j \equiv \text{Prob}\{X_i \in [x_i, x_i + \mathrm{d}x_i) \text{ and } X_j \in [x_j, x_j + \mathrm{d}x_j), \text{ regardless of } X_k\} \quad (72)$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Let (i, j, k) denote any permutation of $(1, 2, 3)$

We define the **conditional density functions**

$$P_i^{(j)}(x_i|x_j)\mathrm{d}x_i \equiv \text{Prob}\{X_i \in [x_i, x_i + \mathrm{d}x_i), \text{ given } X_j = x_j \text{ and regardless of } X_k\} \quad (73)$$

$$P_i^{(j,k)}(x_i|x_j, x_k)\mathrm{d}x_i \equiv \text{Prob}\{X_i \in [x_i, x_i + \mathrm{d}x_i), \text{ given } X_j = x_j \text{ and } X_k = x_k\} \quad (74)$$

$$P_{i,j}^{(k)}(x_i, x_j|x_k)\mathrm{d}x_i\mathrm{d}x_j \equiv \text{Prob}\{X_i \in [x_i, x_i + \mathrm{d}x_i) \text{ and } X_j \in [x_j, x_j + \mathrm{d}x_j), \text{ given } X_k = x_k\} \quad (75)$$

Joint random variables (cont.)

We can analyse the definition of marginal and conditional density function

⇝ They must be non-negative and normalised

The normalisation/closure conditions

$$\int_{-\infty}^{\infty} \mathrm{d}x_i P_i(x_i) = \int_{-\infty}^{\infty} \mathrm{d}x_i P_i^{(j)}(x_i|x_j) = \int_{-\infty}^{\infty} \mathrm{d}x_i P_i^{(j,k)}(x_i|x_j,x_k) = 1 \quad (76)$$

$$\int_{-\infty}^{\infty} \mathrm{d}x_i \int_{-\infty}^{\infty} \mathrm{d}x_j P_{i,j}(x_i,x_j) = \int_{-\infty}^{\infty} \mathrm{d}x_i \int_{-\infty}^{\infty} \mathrm{d}x_j P_{i,j}^{(k)}(x_i,x_j|x_k) = 1 \quad (77)$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

The joint, the marginal and the conditional density functions are interrelated

Consider the probability relations between three outcomes 1, 2 and 3

For (i,j,k) any permutation of $(1,2,3)$, we have

$$p(i \wedge j) = p(i)p(j|i)$$
$$p(1 \wedge 2 \wedge 3) = p(i \wedge j)p(k|i \wedge j)$$
$$p(1 \wedge 2 \wedge 3) = p(i)p(j \wedge k|i)$$

Now, identify outcome i with '$X_i \in [x_i, x_i + dx_i)$'

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

From the definitions of marginal and conditional probabilities, we have

$$P_{i,j}(x_i,x_j)\mathrm{d}x_i\mathrm{d}x_j = P_i(x_i)\mathrm{d}x_i \times P_j^{(i)}(x_j|x_i)\mathrm{d}x_j$$
$$P(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = P_{i,j}(x_i,x_j)\mathrm{d}x_i\mathrm{d}x_j \times P_k^{(i,j)}(x_k|x_i,x_j)\mathrm{d}x_k$$
$$P(x_1,x_2,x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 = P_i(x_i)\mathrm{d}x_i \times P_{j,k}^{(i)}(x_j,x_k|x_i)\mathrm{d}x_j\mathrm{d}x_k$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

We get the equations valid for (i,j,k) any permutation of $(1,2,3)$

$$P_{i,j}(x_i,x_j) = P_i(x_i)P_j^{(i)}(x_j|x_i) \quad (78a)$$
$$P(x_1,x_2,x_3) = P_{i,j}(x_i,x_j)P_k^{(i,j)}(x_k|x_i,x_j) \quad (78b)$$
$$P(x_1,x_2,x_3) = P_i(x_i)P_{j,k}^{(i)}(x_j,x_k|x_i) \quad (78c)$$

The three equations can be complemented by a forth equation[9]

$$P_{i,j}^{(k)}(x_i,x_j|x_k) = P_i^{(k)}(x_i|x_k)P_j^{(i,k)}(x_j|x_i,x_k) \quad (79)$$

[9] From subjecting the arguments giving first equation to condition $X_k = x_k$.

Joint random variables (cont.)

We can derive all subordinate density functions from the joint density

Integrate $P(x_1, x_2, x_3) = P_i(x_i) P^{(i)}_{j,k}(x_j, x_k | x_i)$ over x_j and x_k

$$\rightsquigarrow \quad P_i(x_i) = \int_{-\infty}^{\infty} \mathrm{d}x_j \int_{-\infty}^{\infty} \mathrm{d}x_k (x_1, x_2, x_3) \tag{80}$$

Integrate $P(x_1, x_2, x_3) = P_{i,j}(x_i, x_j) P^{(i,j)}_{k}(x_k | x_i, x_j)$ over x_k

$$\rightsquigarrow \quad P_{i,j}(x_i, x_j) = \int_{-\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3) \tag{81}$$

Joint random variables (cont.)

Each **marginal density function** is obtained by integrating the joint density function over all ignored variables

Joint random variables (cont.)

Divide $P_{i,j}(x_i, x_j) = P_i(x_i) P^{(i)}_{j}(x_j | x_i)$ by $P_i(x_i)$, use $P_i(x_i)$ and $P_{i,j}(x_i, x_j)$

$$\rightsquigarrow \quad P^{(i)}_{j}(x_j | x_i) = \frac{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_{i,j}(x_i, x_j)}}{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_j \int_{\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_i(x_i)}} \tag{82}$$

Divide $P(x_1, x_2, x_3) = P_{i,j}(x_i, x_j) P^{(i,j)}_{k}(x_k | x_i, x_j)$ by $P_{i,j}(x_i, x_j)$, use $P_{i,j}(x_i, x_j)$

$$\rightsquigarrow \quad P^{(i,j)}_{k}(x_k | x_i, x_j) = \frac{P(x_1, x_2, x_3)}{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_{i,j}(x_i, x_j)}} \tag{83}$$

Divide $P(x_1, x_2, x_3) = P_i(x_i) P^{(i)}_{j,k}(x_j, x_k | x_i)$ by $P_i(x_i)$, use $P_i(x_i)$

$$\rightsquigarrow \quad p^{(i)}_{j,k}(x_j, x_k | x_i) = \frac{P(x_1, x_2, x_3)}{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_j \int_{-\infty}^{-\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_i(x_i)}} \tag{84}$$

Joint random variables (cont.)

Each **conditional density function** is obtained by taking the ratio of two integrals of the joint density function

- At the numerator, it is over all ignored variables
- At the denominator, it is over all variables, except conditioning ones

Joint random variables (cont.)

The joint density function determines all marginal and conditional densities

Only some subsets of marginal and conditional densities determine the joint

$$P(x_1, x_2, x_3) = P_{i,j}(x_i, x_j) P_k^{(i,j)}(x_k|x_i, x_j)$$
$$P(x_1, x_2, x_3) = P_i(x_i) P_{j,k}^{(i)}(x_j, x_k|x_i)$$

$P_{i,j}(x_i, x_j) = P_i(x_i) P_j^{(i)}(x_j|x_i)$ into $P(x_1, x_2, x_3) = P_{i,j}(x_i, x_j) P_k^{(i,j)}(x_k|x_i, x_j)$

$$\rightsquigarrow \quad P(x_1, x_2, x_3) = P_i(x_i) P_j^{(k)}(x_j|x_i) P_k^{(i,k)}(x_k|x_i, x_j) \qquad (85)$$

It holds for (i, j, k) denoting any permutation of $(1, 2, 3)$

$\rightsquigarrow$ It is a full conditioning of P

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

We can express each density function as an integral of another densities

Integrate $P(x_1, x_2, x_3) = P_i(x_i) P_{j,k}^{(i)}(x_j, x_k|x_i)$ over x_j, then normalise[10]

$$\rightsquigarrow \quad P_i^{(k)}(x_i|x_k) = \int_{-\infty}^{\infty} \mathrm{d}x_j \, P_{i,j}^{(k)}(x_i, x_j|x_k) \qquad (86)$$

We can permute indexes in $P_{i,j}(x_i, x_j) = P_i(x_i) P_j^{(i)}(x_j|x_i)$, we get

$$P_i(x_i) P_j^{(i)}(x_j|x_i) = P_j^{(x_j)}(x_j) P_i^{(j)}(x_i|x_j)$$

Then, integrate over x_j and normalise

$$\rightsquigarrow \quad P_i(x_i) = \int_{-\infty}^{\infty} \mathrm{d}x_j \, P_j(x_j) P_i^{(j)}(x_i|x_j) \qquad (87)$$

[10] Use $\int_{-\infty}^{\infty} \mathrm{d}x_i \, P_i(x_i) = \int_{-\infty}^{\infty} \mathrm{d}x_i \, P_i^{(j)}(x_i|x_j) = \int_{-\infty}^{\infty} \mathrm{d}x_i \, P_i^{(j,k)}(x_i|x_j, x_k) = 1$.

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Integrate $P(x_1, x_2, x_3) = P_{i,j}(x_i, x_j) P_k^{(i,j)}(x_k|x_i, x_j)$ over x_i and x_j

Then use $P_i(x_i) = \int_{-\infty}^{\infty} \mathrm{d}x_j \int_{-\infty}^{\infty} \mathrm{d}x_k (x_1, x_2, x_3)$

$$\rightsquigarrow \quad P_k(x_k) = \int_{-\infty}^{\infty} \mathrm{d}x_i \int_{-\infty}^{\infty} \mathrm{d}x_j \, P_{i,j}(x_i, x_j) P_k^{(i,j)}(x_k|x_i, x_j) \qquad (88)$$

Integrate $P(x_1, x_2, x_3) = P_i(x_i) P_j^{(k)}(x_j|x_i) P_k^{(i,k)}(x_k|x_i, x_j)$ over x_i

Then use $P_{i,j}(x_i, x_j) = \int_{-\infty}^{\infty} \mathrm{d}x_k \, P(x_1, x_2, x_3)$

$$\rightsquigarrow \quad P_{i,j}(x_i, x_j) = \int_{-\infty}^{\infty} \mathrm{d}x_i \, P_i(x_i) P_j^{(i)}(x_j|x_i) P_k^{(i,j)}(x_k|x_i, x - k) \qquad (89)$$

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Statistical independence of random variables

We define[11] **statistical independence** of RVs

For (i, j, k) any permutation of $(1, 2, 3)$, we have

$$P_i(x_i) = P_i^{(j)}(x_i|x_j) = P_i^{(j,k)}(x_i|x_j, x_k) \qquad (90)$$

A necessary and sufficient condition for X_1, X_2 and X_3 to be statistically independent is that their joint density function $P(x_1, x_2, x_3)$ be equal to the product of their marginal density functions

$$\rightsquigarrow \quad P(x_1, x_2, x_3) = P_i(x_i) P_j(x_j) P_k(x_k) \qquad (91)$$

■

[11] In analogy with the definition of statistical independence of outcomes.

Joint random variables (cont.)

Consider full conditioning $P(x_1, x_2, x_3) = P_i(x_i) P_j^{(k)}(x_j|x_i) P_k^{(i,k)}(x_k|x_i, x_j)$

$$P_i(x_i) = P_i^{(j)}(x_i|x_j) = P_i^{(j,k)}(x_i|x_j, x_k)$$
$$\rightsquigarrow \quad P(x_1, x_2, x_3) = P_i(x_i) P_j(x_j) P_k(x_k)$$

Joint random variables (cont.)

Substitute $P_i(x_i) = P_i^{(j)}(x_i|x_j) = P_i^{(j,k)}(x_i|x_j, x_k)$ into the densities

$$\rightsquigarrow \quad P_j^{(i)}(x_j|x_i) = \frac{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_{i,j}(x_i, x_j)}}{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_j \int_{\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_i(x_i)}}$$

$$\rightsquigarrow \quad P_k^{(i,j)}(x_k|x_i, x_j) = \frac{P(x_1, x_2, x_3)}{\underbrace{\int_{-\infty}^{\infty} \mathrm{d}x_k P(x_1, x_2, x_3)}_{P_{i,j}(x_i, x_j)}}$$

Normalise $P(i)$ to get

$$P(x_1, x_2, x_3) = P_i(x_i) P_j(x_j) P_k(x_k)$$
$$\rightsquigarrow \quad P_i(x_i) = P_i^{(j)}(x_i|x_j) = P_i^{(j,k)}(x_i|x_j, x_k)$$

Joint random variables (cont.)

Average

We define **average** of any three-variate function h

$\rightsquigarrow$ With respect to the set of RVs X_1, X_2 and X_3

$$\langle h(X_1, X_2, X_3) \rangle \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} h\big[x_1^{(n)}, x_2^{(n)}, x_3^{(n)}\big] \qquad (92)$$

$x_j^{(n)}$ is the value found for X_j in the n-th simultaneous sampling of the RVs

We can derive an analytically more convenient form

$$\rightsquigarrow \quad \langle h(X_1, X_2, X_3) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x_1 \int_{-\infty}^{\infty} \mathrm{d}x_2 \int_{-\infty}^{\infty} \mathrm{d}x_3 h(x_1, x_2, x_3) P(x_1, x_2, x_3) \qquad (93)$$

■

Joint random variables (cont.)

$$\langle h(X_1, X_2, X_3) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x_1 \int_{-\infty}^{\infty} \mathrm{d}x_2 \int_{-\infty}^{\infty} \mathrm{d}x_3 h(x_1, x_2, x_3) P(x_1, x_2, x_3)$$

Consider the case in which function h is independent of X_3

We have,

$$\rightsquigarrow \quad \langle h(X_1, X_2) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x_1 \int_{-\infty}^{\infty} \mathrm{d}x_2 h(x_1, x_2) \underbrace{\Big[\int_{-\infty}^{\infty} \mathrm{d}x_3 P(x_1, x_2, x_3)\Big]}_{P_{1,2}(x_1, x_2)}$$

Joint random variables (cont.)

For (i,j) any permutation of $(1,2,3)$, we have the general forms

$$\rightsquigarrow \quad \langle h(X_i, X_j) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x_i \int_{-\infty}^{\infty} \mathrm{d}x_j \, h(x_i, x_j) P_{i,j}(x_i, x_j) \tag{94a}$$

$$\rightsquigarrow \quad \langle h(X_i) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x_i \, h(x_i) P_i(x_i) \tag{94b}$$

Note that $\langle h(X_i) \rangle = \int_{-\infty}^{\infty} \mathrm{d}x_i h(x_i) P_i(x_i)$ has form $\langle h(X) \rangle = \int_{-\infty}^{\infty} h(x) P(x) \mathrm{d}x$

- $\rightsquigarrow$ $\langle X_i^n \rangle$, $\text{mean}\{X_i\}$, $\text{var}\{X_i\}$ and $\text{sdev}\{X_i\}$ can be defined
- $\rightsquigarrow$ (Replace X with X_i)

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Covariance

We define the family of **covariances** of RVs X_i and X_j

$$\text{cov}\{X_i, X_j\} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle \tag{95a}$$

$$= \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \tag{95b}$$

The equivalence is based on the expression of $\langle h(X_i, X_j) \rangle$ and $\langle h(X_i) \rangle$

Compare $\text{cov}\{X_i, X_j\} = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle$ and $\text{var}\{X\} = \langle X^2 \rangle - \langle X \rangle^2$

We have,

$$\text{cov}\{X_i, X_i\} = \text{var}\{X_i\} \tag{96}$$

Covariances are useful quantities regarding sets of joint RVs

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Joint random variables (cont.)

Conditioned averages

We define the family of **conditioned averages** of h

1. With respect to the RVs X_i and X_j
2. Given that $X_j = x_j$

$$\langle h(X_i, X_j | X_j = x_j) \rangle \equiv \int_{-\infty}^{\infty} \mathrm{d}x_i h(x_i, x_j) P_i^{(j)}(x_i | x_j) \tag{97}$$

Conditioned averages are seen as calculated by ignoring all samples in which the value of X_j is not found to be equal (or infinitesimally close) to x_j

- (From the view point of sampling)

$$\langle h(X_1, X_2, X_3) \rangle \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} h\big[x_1^{(n)}, x_2^{(n)}, x_3^{(n)}\big]$$

Conditioned means, conditioned variances, conditioned covariances all follow

Probability and random variable theory
UFC/DC
SA (CK0191)
2018.1
Probability and probability laws
Random variables
General concepts
Moments
Named variables
Discrete variables
Joint variables
Some theorems

Some theorems

Probability and random variable theory

Some theorems

TBD