R403: Probabilistic and Statistical Computations with R

Topic 4: Discrete and continuous univariate and multivariate probability distributions. Special types. Working with probability distributions in R

Kaloyan Ganev

2022/2023

Lecture Contents

Introduction

Introduction

- For a beginning, we will review briefly some basics of probability
- We will then go through some of the probability distributions that are most used in economics
- At the same time, the corresponding R functions will also be discussed

A Brief and Superficial Review of Probability

Sample Space. Events

- Assume that outcomes (e.g. of an experiment) are unpredictable in advance
- Yet, the set of all possible outcomes is known. This set is called the *sample space* (denoted by S)
  Examples: tossing a coin, rolling a die, etc.
- Any subset of S is called an *event*
  Examples: get heads when a coin is tossed; get heads on two coins tossed, etc.

Sample Space. Events (2)

- Union of events:

$$A \cup B; \quad \bigcup_{i=1}^{n} A_i$$

- Intersection of events:

$$A \cap B; \quad \bigcap_{i=1}^{n} A_i$$

- Mutually exclusive events:

$$A \cap B = \emptyset$$

- Complement of an event A:

$$\overline{A} : \text{all elements of } S \text{ that are not in } A$$

Probability Defined on Events

- Let A be any event from S
- Then, probability is a number $P(A)$ defined for A such that
  1. $0 \leq P(A) \leq 1$
  2. $P(S) = 1$
  3. If $A_1, A_2, \ldots$ is a sequence of events and $A_i \cap A_j = \emptyset,\ i \neq j$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- At the intuitive level, the probability of an event is the frequency of its occurrence

Probability Defined on Events (2)

- It is always true that $A \cap \overline{A} = \emptyset$ and $A \cup \overline{A} = S$
- Therefore,
$$P(A \cup \overline{A}) = P(A) + P(\overline{A}) = P(S) = 1$$
- The latter implies
$$P(A) = 1 - P(\overline{A})$$
- Probability of union of two events:[1]
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
- For mutually exclusive events the latter becomes
$$P(A \cup B) = P(A) + P(B)$$

[1] This can be generalized to more than two events but the formula is spared here.

Conditional Probability

- Take events A and B; the probability

$$P(B|A)$$

  is the probability of B given that A already occurred
- In fact, A is a subset of elements of the entire sample set
- Here, as the condition is that specifically A occurred, A itself becomes the sample set
- So in order to asses the conditional probability of B given A, we have to calculate the frequency of $B \cap A (= A \cap B)$ occurring relative to the frequency of A occurring, i.e.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0$$

Conditional Probability (2)

- The latter implies

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$$

- It is also then true that

$$P(B|A) \geq 0$$

$$P(A|A) = 1$$

$$P(\bigcup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A), \quad \text{if } B_1, B_2, \ldots \text{ are mutually exclusive}$$

Independence

- A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

- This implies that A and B are independent if

$$P(B|A) = P(B)$$

- If A and B are not independent, they are called *dependent*

The Bayes Formula

- From $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$ follows that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad P(B) \neq 0 \quad \text{(Bayes formula)}$$

- Law of total probability: if $\{B_i : \; i = 1, 2, 3, \ldots\}$ is a finite or countable partition of a sample space (i.e. all B_i are pairwise disjoint), then if A is on the same probability space,

$$P(A) = \sum_i P(A|B_i)P(B_i)$$

- Using this, the Bayes formula can be written also as

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

Random Variables

Random Variables

- Random variables are functions that map events from a sample space to real numbers
- Indicator random variable:

$$I = \begin{cases} 1, & \text{if a property is present} \\ 0, & \text{if a property is absent} \end{cases}$$

- **Discrete** random variable: can take on a finite or countable number of values
- **Continuous** random variable: can take on a continuum of values

Random Variables (2)

- Cumulative distribution function (cdf):

$$F(x) = P(X \leq x)$$

- Properties of cdf:
  - $F(x)$ is non-decreasing
  - $\lim_{x \to \infty} F(x) = F(\infty) = 1$
  - $\lim_{x \to -\infty} F(x) = F(-\infty) = 0$
- Implication:

$$P(a < X \leq b) = F(b) - F(a)$$

- Quantiles: values of the random variable that correspond to a given value of the cdf[2]

[2] You can think of the quantile function as of the inverse of the cdf.

Random Variables (3)

- The term *cdf* is often used interchangeably with the term *probability distribution*
- Both in fact stand for a mathematical model of probability, i.e. the rule by which we assign probabilities to values (events)
- Knowledge on distributions is used in hypotheses testing, for example

Expectation of a Random Variable

- Discrete:

$$\mathsf{E}(X) = \sum_{x:\ p(x)>0} x \cdot p(x)$$

- Continuous:

$$\mathsf{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

- Let $g(\cdot)$ be a real-valued function; then:

$$\mathsf{E}(g(X)) = \sum_{x:\ p(x)>0} g(x) \cdot p(x)$$

$$\mathsf{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$$

Expectation of a Random Variable (2)

- $\mathsf{E}(X)$ is also called the *mean* of a random variable
- $\mathsf{E}(X^n)$ is called the *n*th *moment* of a random variable
- If a and b are constants,

$$\mathsf{E}(aX + b) = a\,\mathsf{E}(X) + b$$

- Variance:

$$\mathsf{Var}(X) = \mathsf{E}[X - \mathsf{E}(X)]^2 = \mathsf{E}(X^2) - [E(X)]^2$$

- An important relationship

$$\mathsf{Var}(X) = \mathsf{E}[\mathsf{Var}(X|Y)] + \mathsf{Var}[\mathsf{E}(X|Y)]$$

- Law of iterated expectations:

$$\mathsf{E}(X) = \mathsf{E}[\mathsf{E}(X|Y)]$$

Discrete Random Variables

- Probability mass function (pmf):

$$p(x) = P(X = x)$$

- The pmf is positive for a countable number of values of X; for the remaining ones it is zero
- Property:

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

- cdf:

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

Bernoulli Random Variables

- Characterize experiments in which the outcome is classified either as a *success* (1) or as a *failure* (0)
  Example: tossing a coin
- pmf (for $p \in (0,1)$):

$$\begin{aligned} p(0) &= P(X=0) = 1-p \\ p(1) &= P(X=1) = p \end{aligned}$$

  Can be written also as

$$p(x) = p^x (1-p)^{1-x}$$

- cdf:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \le x < 1 \\ p, & x \ge 1 \end{cases}$$

Bernoulli Random Variables (2)

- Mean:

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$

- Variance:

$$\begin{aligned} \mathrm{Var}(X) &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = \\ &= p^2 \cdot (1-p) + (1-p)^2 \cdot p = \\ &= (1-p) \cdot (p^2 + p - p^2) = p \cdot (1-p) \end{aligned}$$

- Bernoulli random variables are a special case of the binomial ones
- Therefore we will postpone R simulation for later

Binomial Random Variables

- The binomial distribution shows the number of successes in n *independent*[3] Bernoulli trials
- Specifically, the probability of x successes in n trials (the pmf) is:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- cdf:

$$F(x) = P(X \leq x) = \sum_{i=1}^{x} \binom{n}{i} p^i (1-p)^{n-i}$$

- Mean: np
- Variance: $np(1-p)$

[3] If the experiment implies sampling, then there is replacement of elements.

Binomial Random Variables (2)

- Example problem: Find the probability of getting heads two times out of ten fair coin tosses
- Obviously, we have to use the pmf to solve it
- We will go to R for the solution:

```
dbinom(2, 10, 0.5, log = FALSE)
```

- Example problem: What is the probability of not getting a pass (60%) if you answer randomly at an exam in which there are 20 multiple-choice questions (the number of options is 4)?
- Solution using the cdf:

```
pbinom(11, 20, 0.25, lower.tail = TRUE, log.p = FALSE)
```

Binomial Random Variables (3)

- Example problem: find the number of correct answers so that your test score is in the 75th percentile
- Solution using the quantile function:

```
qbinom(0.75, 20, 0.25, lower.tail = TRUE, log.p = FALSE)
```

- Generate 100 random numbers from the binomial distribution, where $p = 0.6$ and $n = 20$:

```
rbinom(100, 20, 0.6)
```

- Plot them:

```
plot(rbinom(100, 20, 0.6), type = "p")
```

- Histogram:

```
hist(rbinom(100, 20, 0.6), breaks = 20)
```

Geometric Random Variables

- Assume that in independent Bernoulli trials, the probability of success is p
- Problem to solve: What is the probability of getting x failures before a success occurs?
- The random variable that measures that probability is the *geometric* random variable
- pmf:

$$p(x) = P(X = x) = (1 - p)^x p, \quad x = 0, 1, 2, \ldots$$

- Example problem: What is the probability of tossing 5 times a fair coin and getting tails before we get heads on the sixth toss?

```r
dgeom(5, 0.5, log=FALSE)
```

- Problem to do: Prove that $\sum_{x=0}^{\infty} p(x) = 1$

Geometric Random Variables (2)

- The cdf is:

$$F(x) = 1 - (1-p)^{x+1} \quad x = 1, 2, \ldots$$

- Example of usage: what is the probability of tossing a coin between zero and five times before getting heads?

```
pgeom(5, 0.5, lower.tail = TRUE, log.p = FALSE)
```

- Quantiles and random number generation:

```
qgeom, rgeom
```

- Mean: $\dfrac{1-p}{p}$
- Variance: $\dfrac{1-p}{p^2}$

Hypergeometric Random Variables

- The hypergeometric distribution is another special case of the binomial distribution
- Difference: There is no replacement therefore the condition of independence is violated
- The pmf is (measures the probability of k successes):

$$p(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

where N is the size of the population, K is the total number of successes in the population.

- cdf: too long to write here :)

Hypergeometric Random Variables (2)

- Mean: $n \cdot \dfrac{K}{N}$
- Variance: $n \cdot \dfrac{K}{N} \cdot \dfrac{N-K}{N} \cdot \dfrac{N-n}{N-1}$
- The hypergeometric distribution in R:

```
dhyper, phyper, qhyper, rhyper
```

Poisson Random Variables

- Measures the number of events that occur for a fixed interval of time/space
- Main assumptions:
  - Each event occurs independently of the time since the last event occurred
  - The mean rate of occurrence per unit of time is constant
- Examples: customers per hour in a shop; accidents per day in a city; raindrops per square centimetre; etc.
- pmf:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where $\lambda = \mathrm{E}(X) = \mathrm{Var}(X)$ is the constant mean rate of occurrence

Poisson Random Variables (2)

- cdf: also not too pleasant to write, so it's skipped here
- Example: 15 cars arrive at a gas station on average per hour. What is the probability that 10 cars arrive during a chosen hour?

$$p(10) = P(X = 10) = \frac{15^{10}e^{-15}}{10!} \approx 4.9\%$$

(Obviously, here $\lambda = 15$)

Poisson Random Variables (3)

- Calculate the same problem in R:

```
dpois(10, 15, log = FALSE)
```

- Calculate the probability of having 10 or less cars arrive:

```
ppois(10, 15, lower.tail = TRUE, log.p = FALSE)
```

- Quantiles and random number generation:

```
qpois, rpois
```

Continuous Random Variables

- The set of possible values is *uncountable*
- Probability density function (pdf): a non-negative function defined for all $x \in (-\infty, \infty)$
- Denoted by $f(x)$, has the property that for any $B \subset \mathbb{R}$,

$$P(X \in B) = \int_B f(x)\,dx$$

- If $B = (-\infty, x]$, then

$$P(X \in B) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = F(x)$$

is in fact the cdf

Continuous Random Variables (2)

- It is clear that
$$F'(x) = f(x)$$
- If $B = [a, b]$ then
$$P(X \in B) = \int_a^b f(x)\,dx = F(b) - F(a)$$
- Letting $a = b$ leads to
$$P(X \in B) = \int_a^a f(x)\,dx = F(a) - F(a) = 0$$

In other words, the probability of getting a specific value out of infinitely many is zero

Continuous Random Variables (3)

- If $B = \mathbb{R}$, then

$$P(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

  The latter implies $F(\infty) = 1$ and $F(-\infty) = 0$

- Given that $\int_{a}^{a} f(x)\,dx = 0$,

$$P(X \leq x) = P(X < x)$$

Uniform Random Variables

- For the interval $[a, b]$, the pdf is:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & elsewhere \end{cases}$$

- The cdf is:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & x \in [a, b) \\ 1, & x \geq b \end{cases}$$

Uniform Random Variables (2)

- Mean:

$$\mathsf{E}(X) = \frac{1}{2}(a + b)$$

- Variance:

$$\mathsf{Var}(X) = \frac{1}{12}(b - a)^2$$

Uniform Random Variables (3)

- To calculate density at a given value in R:

```
dunif(0.7, min = 0, max = 1, log = FALSE)
```

- The minimum and the maximum value of the argument are 0 and 1 respectively by default, so the latter is equivalent to:

```
dunif(0.7)
```

- To plot the pdf:

```
plot(dunif,0,1,main = "The Uniform pdf")
```

- To calculate the cdf.:

```
punif(0.7, min = 0, max = 1, lower.tail = TRUE, log.p = FALSE)
```

Uniform Random Variables (3)

- To plot the cdf:

```
plot(punif, 0, 1, main = "Uniform cdf")
```

- Calculate median:

```
qunif(0.5, min = 0, max = 1, log = FALSE)
```

- Generate 10000 values of a uniformly distributed random variable:

```
runif(10000, min = 0, max = 1)
```

- And plot them:

```
plot(runif(10000, min = 0, max = 1), type = "l", col = "red",
     main = "Uniform RV")
```

- Histogram:

```
hist(runif(10000, min = 0, max = 1), breaks = 50, col = "red",
     main = "Uniform RV (histogram)")
```

The Poisson Process

- Homogeneous case: the process counts events that occur at a constant rate in time:

$$P[N(t+h) - N(t) = k] = \frac{e^{-\lambda h}(\lambda h)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

- Non-homogeneous case: the rate of occurrence is not constant; we do not deal with this case

Exponential Random Variables

- The continuous analogue of geometric random variables
- Model the time between events in a Poisson process
- The pdf is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- λ is the mean rate at which events occur, while $\dfrac{1}{\lambda}$ is the mean waiting time between two events

Exponential Random Variables (2)

- Mean:

$$\mathsf{E}(X) = \frac{1}{\lambda}$$

- Variance:

$$\mathsf{Var}(X) = \frac{1}{\lambda^2}$$

- Note: The exponential distribution is a special case of the gamma distribution (along with the Erlang and the χ^2 distributions)

Exponential Random Variables (3)

- Implementation in R:

```
dexp(2, 0.25)
```

- Plot the pdf:

```
x <- seq(0,20, by = 0.01)
y <- dexp(x,0.25)
plot(x,y,type = "l")
```

- The c.d.f. is:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- Plot the cdf

```
y <- pexp(x, 0.25)
plot(x, y, type = "l")
```

Exponential Random Variables (4)

- Example: Let x be the amount of time that a customer is processed by a cashier at a department store. What is the probability that a customer is processed between 3 and 4 minutes if on average a customer is processed 4 minutes?
- Solution: We have to find $P(3 < X < 4)$. This is accomplished by:

```
pexp(4, 0.25) - pexp(3,0.25)
```

- Next, find the time for which half of the customers are being processed

```
qexp(0.5, 0.25)
```

i. e. we found the median time

The Gamma Function

- An extension of the factorial ($n!$) function
- For positive integers:

$$\Gamma(n) = (n-1)!$$

- For complex numbers with a positive real part:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx, \quad Re(z) > 0$$

- Not defined for negative integers or complex numbers with negative real part, and also for zero

The Gamma Function: Implementation in R

- Through the `factorial()` function
- Example:

```
x <- seq(1, 10, by=.1)
y <- factorial(x)
plot(x, y, type = "l")
```

Gamma Random Variables

- pdf of Gamma distribution with parameters $\lambda > 0$, $\alpha > 0$:

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- λ is called the *rate* parameter, while α is called the *shape* parameter
- (Check what happens with the pdf if $\alpha = 1$, for example)
- How does it differ from the exponential distribution? It models waiting time for 1 *or more* events ahead[4]

[4] When only the waiting time to the next event is modelled, this is the case of the exponential distribution.

Gamma Random Variables (2)

- Mean:

$$\mathsf{E}(X) = \frac{\alpha}{\lambda}$$

- Variance

$$\mathsf{Var}(X) = \frac{\alpha}{\lambda^2}$$

Normal Random Variables

- The pdf of the normal distribution is given by the formula:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in (-\infty, \infty)$$

- μ and σ^2 are correspondingly the mean and the variance of the normal distribution
- Those two parameters uniquely define each normal distribution
- If $\mu = 0$ and $\sigma^2 = 1$, then the distribution becomes the *standard normal distribution*
- If $Y = a + bX$, then

$$\mathsf{E}(Y) = \mathsf{E}(a + bX) = a + b\mu$$

$$\mathsf{Var}(Y) = \mathsf{Var}(a + bX) = b^2\sigma^2$$

Normal Random Variables (2)

- To see the density at a given value (i. e. the height of the pdf at that value) – in this case we take as an example the density at 5:

```
dnorm(5, mean = 0, sd = 1, log = FALSE)
```

- In the case of the standard normal pdf, you can issue just:

```
dnorm(5)
```

- You can plot the pdf with:

```
plot(dnorm,-3,3, main = "Standard normal distribution")
```

- What if you want to plot a 'non-standard' normal distribution?

```
x <- seq(-20, 20, by=0.1)
y <- dnorm(x, mean = 2, sd = 5)
plot(x, y, type = "l")
```

Normal Random Variables (3)

- The cdf is the integral of the pdf from $-\infty$ to x
- To calculate the cdf:

```
pnorm(5, mean = 2, sd = 5, lower.tail = TRUE, log.p = FALSE)
```

- For example, to find the probability that a value is within 1.96 standard deviations from the mean:

```
pnorm(1.96, lower.tail = TRUE) - pnorm(-1.96, lower.tail = TRUE)
```

- To plot the normal cdf:

```
plot(pnorm, -5, 5, main = "Normal cdf")
```

Normal Random Variables (4)

- The quantile function is the inverse of the distribution – it gives the value x at which $P(X \leq x) = p$, where p is pre-specified
- It is implemented via:

```
qnorm(0.25, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
```

- For example, this can be used to find the median of the standard normal distribution:

```
qnorm(0.5)
```

- ...or to find the lower and upper bounds of a 95% confidence interval:

```
c(qnorm(0.025),qnorm(0.975))
```

Normal Random Variables (5)

- To generate a normal random variable with 10000 values:

```
rnorm(10000, mean = 2, sd = 5)
```

- To plot such a variable:

```
plot(rnorm(10000, mean = 2, sd = 5), type = "l", col = "blue",
    main = "White noise")
```

- To plot its histogram:

```
hist(rnorm(10000, mean = 2, sd = 5), breaks = 50, col = "orange")
```

The χ^2 Distribution

- Let $X_1, \ldots, X_k$ be k independent standard normal random variables. Then:

$$(X_1^2 + \ldots + X_k^2) \sim \chi^2(k)$$

- Read as “chi-square distribution with k degrees of freedom”
- Main usage: for hypotheses testing or for construction of confidence intervals
- Note: We are considering the so-called centred chi-square distribution; it is generalized in its non-centred version

The χ^2 Distribution (2)

- The pdf is:

$$f(x,k) = \begin{cases} \dfrac{x^{\frac{(k-2)}{2}} e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

- Note that the χ^2 distribution is a special case of the gamma distribution where $t = \dfrac{k}{2}$ and $\lambda = \dfrac{1}{2}$
- R implementation:

```
dchisq(3, 5)
```

- Plot the pdf:

```
x <- seq(0, 20, by = 0.01)
y <- dchisq(x, 5)
plot(x, y, type="l")
```

The χ^2 Distribution (3)

- The cdf is:

$$F(x,k) = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{2^{k/2}\Gamma\left(\frac{k}{2}\right)}$$

where γ is the lower incomplete gamma function defined as:

$$\gamma(s,x) = \int_0^x t^{s-1}e^t dt$$

- Implementation in R:

```
pchisq(3, 5)
```

- Plot the cdf:

```
x <- seq(0, 20, by = 0.01)
y <- pchisq(x, 5)
plot(x, y, type = "l")
```

- Quantiles and random numbers:

```
qchisq, rchisq
```

The Beta Function

- Known also as 'the Euler integral'
- Defined as:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$$

where x and y are complex numbers having positive real parts.

The F Distribution

- If $X_1 \sim \chi^2(k_1)$ and $X_2 \sim \chi^2(k_2)$, then $\dfrac{X_1/k_1}{X_2/k_2} \sim F(k_1, k_2)$
- Used mainly in hypotheses testing
- pdf:

$$f(x, k_1, k_2) = \sqrt{\frac{(k_1 x)^{k_1}\, k_2^{k_2}}{xB\left(\frac{k_1}{2}, \frac{k_2}{2}\right)}}$$

- R implementation:

```
df(3, 2, 5)
```

The F Distribution (2)

- The c.d.f.:

$$F(x, k_1, k_2) = I_{\frac{k_1 x}{k_1 x + k_2}}\left(\frac{k_1}{2}, \frac{k_2}{2}\right)$$

  where I is the regularized incomplete beta function (we will leave these complications to your own curiosity)

- In R:

```
pf(3, 2, 5)
```

- Quantiles, random variables:

```
qf, rf
```

The t Distribution

- Arises in estimating the mean of a normally distributed variable when sample size is small and the population variance is unknown
- Used in statistical tests, linear regression analysis, confidence interval construction, etc.

The t Distribution (2)

- Consider the random variable x. Let the sample size is n and sampling is from a normally distributed population with population mean μ
- Let the sample mean of the random variable be:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- and let the sample variance be:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

The t Distribution (3)

- The statistic:
$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$
has t distribution with $n - 1$ degrees of freedom
- Interpreted as the distribution of the difference between the population mean and the sample mean divided by the sample standard deviation (the latter normalized by $\sqrt{n}$)
- Continuous distribution, symmetrical
- The pdf is:
$$f(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{(n-1)\pi}\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$$

The t Distribution (4)

- R implementation:

```
dt(0, 15)
```

- Plot the pdf:

```
x <- seq(-20, 20, by = 0.01)
y = dt(x, 99)
plot(x, y, type = "l")
```

- The c.d.f.: again too long to write; implementation in R:

```
pt(0, 15)
```

- Quantiles and random number generation:

```
qt, rt
```

Multivariate Distributions

What is a Multivariate Distribution?

- Similarly to the univariate case, a multivariate probability distribution is a model of probability
- However, this model describes more than one random variable at a time
- In other words, instead of having a scalar random variable, we work with a random vector
- A synonym of *multivariate distribution* is *joint distribution*
- The latter is maybe more informative since it captures the potential interaction of random variables

The Joint Probability Distribution

- Take the random vector $(X_1, X_2, \ldots, X_n)$
- The joint probability distribution gives the probability that each of the random variables takes a specific value or falls in a specified range
- Joint probability distributions are expressed through:
  - Their joint cumulative distribution function (cdf)
  - Their joint probability density function (pdf)/probability mass function (pmf)

The Joint cdf, pdf, and pmf

- Joint cdf:

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

- Joint pmf (discrete case):

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

- Joint pdf (continuous case)

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}$$

Example: Multivariate Normal

- pdf:

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}}$$

  where $\mathbf{X} = (X_1, X_2, \ldots, X_n)$
- No analogical capabilities in base R
- The **mvtnorm** package: for working with multivariate t and normal distributions

R Example: Bivariate Normal

- Calculate density:

```
library(mvtnorm)

sigmamat <- matrix(c(1, 0.6, 0.6, 1), nrow = 2)
sigmamat

dmvnorm(x = c(1,2), mean = c(0.05, -0.05), sigma = sigmamat)
```

- Cumulative:

```
pmvnorm(lower = -Inf, upper = c(1,2), mean = c(0.05, -0.05),
    sigma = sigmamat)
```

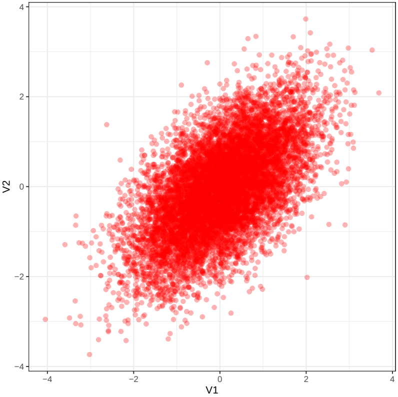
R Example: Bivariate Normal (2)

- Random number generation, correlated variables:

```
rvec1 <- rmvnorm(n = 10000, mean = c(0.05, -0.05), sigma =
    sigmamat)
rvec1 <- as.data.frame(rvec1)
```

- See the following slide for a picture

R Example: Bivariate Normal (3)



R Example: Bivariate Normal (4)

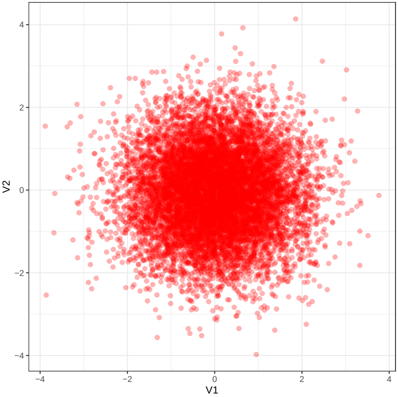
- Same but if the two variables are uncorrelated

```
sigmamat2 <- matrix(c(1, 0, 0, 1), nrow = 2)
sigmamat2

rvec2 <- rmvnorm(n = 10000, mean = c(0.05, -0.05), sigma =
    sigmamat2)
rvec2 <- as.data.frame(rvec2)
```

- See next picture

R Example: Bivariate Normal (5)



References

- Gamma Distribution — Intuition, Derivation, and Examples
- Hogg, R., McKean, J., and A. Craig (2018): *Introduction to Mathematical Statistics*, Pearson, 8th ed.
- Krishnamoorty, K. (2016): *Handbook of Statistical Distributions with Applications*, CRC Press, 2nd ed.
- Pinsky, M., and S. Karlin (2010): *An Introduction to Stochastic Modelling*, Academic Press, fourth ed.
- Ross, S. (2014): *A First Course in Probability*, Peason, 9th ed.
- Ross, S. (2019): *Introduction to Probability Models*, Academic Press, 12th ed.