# Continuous Distributions

## Lecture 9

Termeh Shafie



## recall from last lecture: discrete random variables

A random variable is discrete if its range is a countable (finite or infinite) set.

If $X$ is a discrete random variable, the function given by $f(x) = P(X = x)$ for each $x$ within the range of $X$ is called the probability distribution of $X$, also called probability mass function (pmf)

the probability of an event $A$ associated with a discrete random variable $X$ is found by summing up its probability mass function over the values in that set: $P(X \in A) = \sum_{x \in A} f(x)$

this is not feasible when finding the probability of an event $A$ associated with a continuous random variable $X$



## continuous random variables

A continuous random variable is one that takes values over a continuous range: the whole real line; an interval on the real line, perhaps infinite; or a disjoint union of such intervals.

+

A continuous random variable $X$ must have the property that no possible value has positive probability: $P(X = x) = 0$ for all $x \in \mathbb{R}$



## probability density function

A random variable $X$ is continuous if there is a nonnegative function $f(x)$, called the probability density function (pdf) of $X$, such that

$$P(X \in A) = \int_A f(x)dx$$

for every subset of $A$ of the real line. Specifically, the probability that $X$ is in an interval is

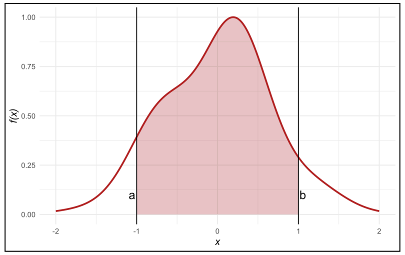
$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

For any PDF we know that $f(x) \geq 0$ for all values of $x$ and the total area under the whole graph is 1

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Note: $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < x < b)$

# probability density function




For $f(x)$ to be a legitimate pdf, it must satisfy the following two conditions:

1. $f(x) \geq 0$ for all values of $x$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ i.e. area under the entire graph of $f(x) = 1$



# probability density function

*exercise 1*

Let $X$ be a continuous random variable with probability density function $f(x) = 3x^2, \quad 0 \leq x \leq 1$

(a) Verify that $f(x)$ is a valid probability function

(b) What is $P\,(1/2 \leq X \leq 1)$?

(c) What is $P(X = 1/2)$?

*exercise 2*

Let $X$ be a continuous random variable with probability density function $f(x) = \dfrac{x^3}{4}$ for $0 \leq X \leq c$.

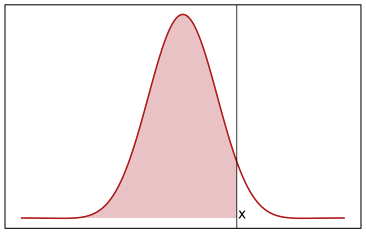
What is the value of the constant $c$ that makes $f(x)$ a valid probability density function?



# cumulative distribution function

For a continuous random variable $X$ with pdf $f(x)$ its cumulative distribution function (cdf) is defined as follows

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$





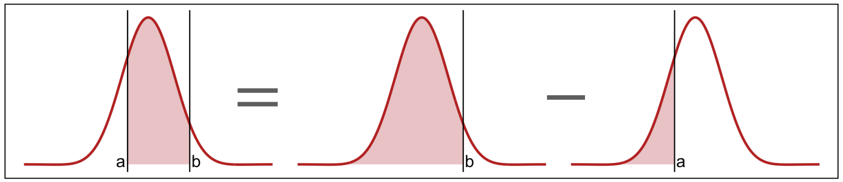

# computing probabilities with cdf

Let $X$ be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Then for any value $a$ we have that

$$P(X \leq a) = F(a) \qquad P(X > a) = 1 - F(a)$$

and for any two values $a < b$

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



## computing probabilities with cdf

*exercise 3*

Random variable $T$ is distributed with the following probability density function:

$$f(t) = \begin{cases} ct(t-1) & \text{for } 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Calculate the value of $c$.

(b) Calculate the cumulative distribution function $F(t)$.

(c) Use the cdf $F(t)$ to calculate $P(1/3 \le T \le 2/3)$.



## summary: pdf and cdf

**PDF of a continuous random variable $X$**

- consider an integral – continuous analogue to 'sums'
- no area in a line – so no probability assigned to RV taking on a specific value

$f(x) \ge 0, \quad$ for all $x \in \mathbb{R}$

$f$ is piecewise continuous

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

**CDF of a continuous random variable $X$**

- CDF is found by integrating the PDF
- PDF is found by differentiating the CDF
- the CDF is always non-decreasing

$$F(x) = \int_{-\infty}^{x} f(y)\,dy \quad \text{for } -\infty < x < \infty$$

Why do we use $y$ instead of $x$?



## expected value of a continuous random variable

Let $X$ be a continuous random variable with pdf $f(x)$. The expected value $E(X)$ is calculated as a weighted integral

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Let $X$ be a continuous random variable with pdf $f(x)$. If $h(X)$ is any real-valued function of $X$ then we can calculate an expected value for that as

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



## variance of a continuous random variable

Let $X$ be a continuous random variable with pdf $f(x)$ and mean $E(X) = \mu$. The variance $V(X)$ is is the expected value of the squared distance to the mean

$$V(X) = E((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu^2)\cdot f(x)dx$$

The standard deviation is given by $\sqrt{V(X)}$.

# theoretical joint distributions

For two continuous random variables, we can write their joint pdf the same way: $f(x, y)$
"summing" the small bits of probability $f(x, y)dxdy$ over some region $X \in A, Y \in B$

Let $X, Y$ be a continuous random variables. The joint pdf for $X$ and $Y$ is $f(x, y) \geq 0$

The joint range is the set of pairs $(x, y)$ that have non-zero density.

The double integral over all values must be 1

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$$



# theoretical joint distributions

*exercise 4*

Let $X$ and $Y$ be two jointly continuous random variables with the following joint pdf

$$f(x, y) = \begin{cases} x + cy^2 & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find a sketch the joint range of $X$ and $Y$ (i.e. $\Omega_{X,Y}$).

(b) Find the constant $c$ that makes $f(x, y)$ a valid joint pdf.

(c) Find $P(0 \leq X \leq 1/2,\ 0 \leq Y \leq 1/2)$.



# marginal distributions

Let $X, Y$ be jointly distributed continuous random variables with joint pdf $f(x, y)$.
The marginal pdf's of $X$ and $Y$ are respectively given by the following:

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

$$f(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Note this is exactly like for joint discrete random variables, with integrals instead of sums.



# theoretical joint distributions

*exercise 5*

Find the marginal pdf $f(x)$ and $f(y)$ given the joint pdf:

$$f(x, y) = \begin{cases} x + \frac{3}{2}y^2 & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

# some continuous random variables and their pdfs

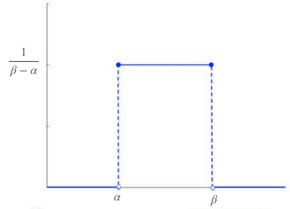
## uniform distribution $X \sim \text{Unif}(\alpha, \beta)$

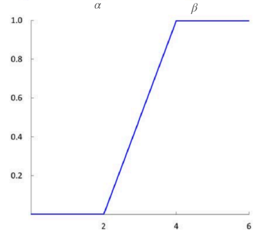
A continuous random variable $X$ has uniform distribution on the interval $[\alpha, \beta]$ for values $\alpha \leqslant \beta$ if it has the following pdf:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{for } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

The cdf is given by

$$f(x \mid \alpha, \beta) = \begin{cases} 0 & \text{if } x < a \\ \frac{x - \alpha}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 1 & \text{if } x > \beta \end{cases}$$






# some continuous random variables and their pdfs

## normal distribution $X \sim N(\mu, \sigma^2)$

A continuous random variable $X$ has normal distribution with parameters $\mu$ and $\sigma^2$ if it has the following pdf:

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

If continuous random variable $X \sim N(\mu, \sigma^2)$ then random variable $Z$ defined as

$$Z = \frac{X - \mu}{\sigma}$$

z scores

has standard normal distribution $Z \sim N(0,1)$



# some continuous random variables and their pdfs

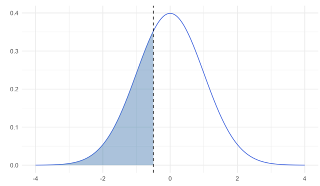
## standard normal distribution $Z \sim N(0,1)$

The normal distribution with parameters $\mu = 0$ and $\sigma = 1$ is the standard normal distribution and a random variable with that distribution is a standard normal random variable, usually named $Z$ and with the following probability density function.

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

The corresponding cumulative distribution function is written $\Phi(z)$

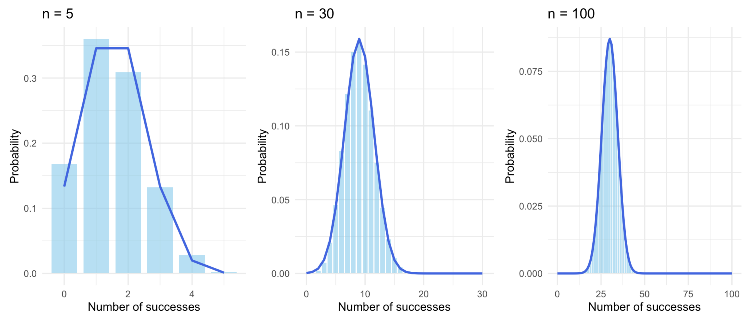
$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}}\, dx$$





# some continuous random variables and their pdfs

## the importance of normal distribution…




$np \geq 10$

## central limit theorem (CLT)

Regardless of the shape of a population's distribution, if the sample size $n$ is large enough ($n \geq 30$) and there is finite variance, then…

- the distribution of the sample means will be approx. normal
  → shape of distribution of $X$ becomes more bell-shaped and symmetric
- centre of the distribution of $\overline{X}$ remains $\mu$
- the spread of the distribution increases and it becomes more 'peaked'



## law of large numbers (LLN)

The law of large numbers states that for an increasing number of trials, the sample average should approach the population average.

- guarantee for long-term and stable results of random events
- necessary for statistical modelling → remember asymptotic normality, efficiency and consistency?
- the spread of the distribution increases and it becomes more 'peaked'

Exceptions: samples from Cauchy and some Pareto distributions ($\alpha < 1$) may not converge as n increases! → often due to heavy tails (skewness)



## some continuous random variables and their pdfs

**gamma distribution $X \sim$ Gamma$(\alpha, \beta)$**

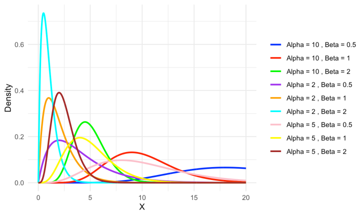
A continuous random variable $X$ has Gamma distribution with parameters $\alpha$ and $\beta$ (both positive) if

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{x/\beta} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma(\alpha)$ is the gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx,$$

which cannot be expressed in closed form analytical solution.




$$\alpha = 1 \implies \text{Exponential} \left(\beta = \frac{1}{\lambda}\right)$$

$$\alpha = \frac{v}{2}, \beta = 2 \implies \text{chi-square} \quad \chi^2(v)$$



## some continuous random variables and their pdfs

**exponential distribution $X \sim$ Exp$(\lambda)$**

A continuous random variable $X$ has exponential distribution with parameter $\lambda$, for some $\lambda > 0$, if it has the following pdf

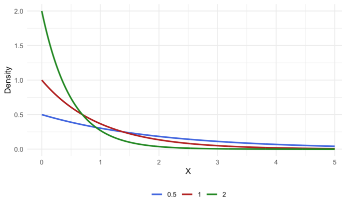
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

and the following cdf

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

This distribution is **memoryless** i.e.

$$P(X \geq a \,|\, X \geq b) = P(X \geq a - b)$$




The exponential distribution is a specific version of gamma family distributions…

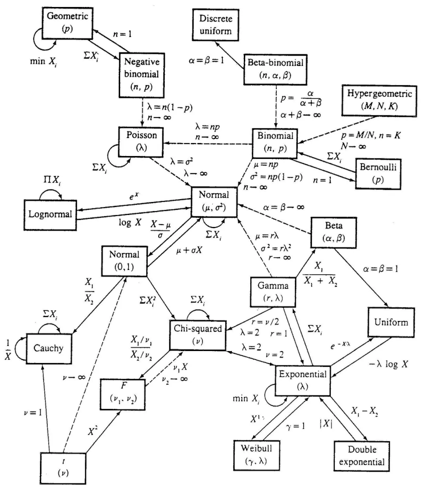
# some continuous random variables and their pdfs

## important distributions for statistical hypothesis tests

- Chi-squared ($\chi^2$) distribution (special case of the gamma distribution, where $\alpha = n/2$ and $\beta = 2$)
- The (Student's) *t* Distribution (like normal but "thicker" tails)
- The F Distribution (ratio of two $\chi^2$ distributed variables)

read up on these distributions (and others) in your text book






https://www.math.wm.edu/~leemis/chart/UDR/UDR.html