37161 Probability and Random Variables

Lecture 5

Continuous Random Variables

- So far, we have only looked at discrete random variables. That is, variables which can only take a value from a (possibly finite, possibly infinite) sample space and not any values in between these.

- For example, if we want to know how long (in minutes) it will be until the next train arrives, assuming one train arrives every ten minutes, then our sample space for this is a continuous interval [0, 10) .

- As this interval contains an uncountable number of points, we cannot write down a probability mass function of the form $P(X = k) = \ldots$

Continuous Random Variables

- In any case, the probability that the random variable takes any given value exactly is infinitesimally small. Even if we record that the train arrived after 7.5 minutes, in reality, it is exceptionally unlikely that it arrived at this time exactly and not, say, 7.500000000100001030400000000007 minutes.

- Whatever degree of precision we use, the probability of getting an exact value is ≈ 0.

- Instead of a probability mass function, we define a **probability density function** for the continuous variable X to be $f(x)$ such that $P(a < X < b) = \int_a^b f(x)dx$.

Probability Density Functions

- The probability density function $f(x)$ gives a relative measure of how likely the random variable is to take a value in a given region. It is not, though, itself a probability.

- We do still have some standard properties of a density function:

- $\int_{-\infty}^{\infty} f(x)dx = 1$ since the probability of an event from somewhere in the sample space must be 1.

- $f(x) \geq 0$ or else it would be possible to have a region $[c,d]$ such that

  $$P(c < X < d) = \int_{c}^{d} f(x)dx < 0$$

Probability Density Functions

- Note, though, that we do not need $f(x) \leq 1$.
- We do still require $\int_{-\infty}^{\infty} f(x)dx = 1$ but the probability density function can exceed 1 on short intervals, provided the integral over that interval stays no greater than 1.
- An intuitive interpretation of the density function is that, for very small $\varepsilon > 0$,

$$P\left(a - \frac{\varepsilon}{2} < X < a + \frac{\varepsilon}{2}\right) = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f(x)dx \approx \varepsilon f(a)$$

- In other words, for very small $\varepsilon > 0$, the probability that $X = a$ (with a margin of error no more than ε centred around a is approximately $\varepsilon f(a)$.

Expectation of Continuous Random Variables

- We cannot, of course, calculate the expectation of a continuous random variable as we did for discrete variables, since the probability mass function $P(X = k)$ is not defined.

- Just as to calculate the probabilities, we integrated (rather than summing all outcomes for an event), we again can integrate to obtain equivalent values to $E(X) = \sum k \times P(X = k)$.

- For continuous variables, we instead have, for continuous random variable X with density function $f(x)$, $E(X) = \int_{-\infty}^{\infty} x f(x) dx$.

Expectation of Continuous Random Variables

- Similarly, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$, $E(X-10) = \int_{-\infty}^{\infty} (x-10)f(x)dx$ etc.

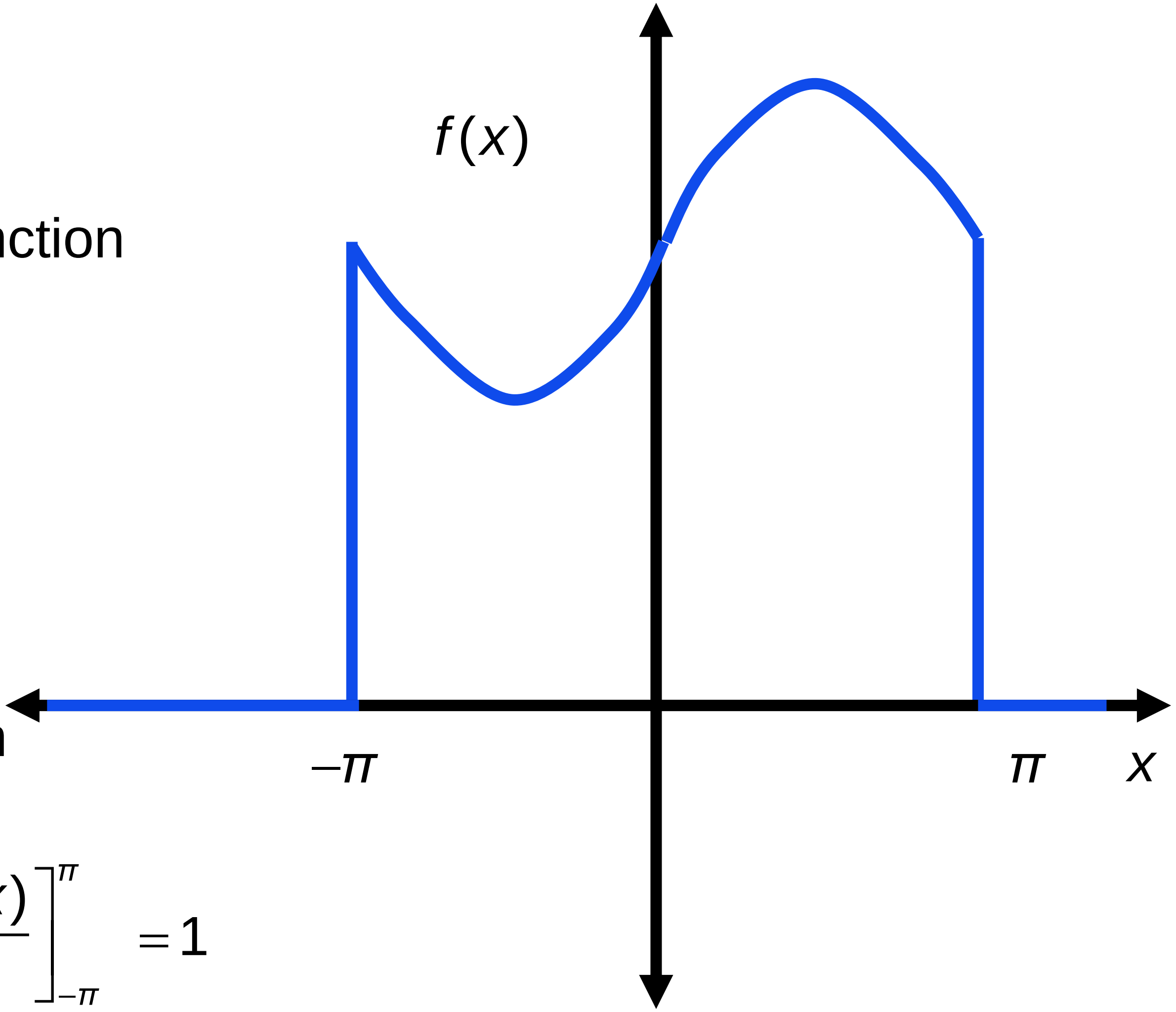
- We can therefore calculate the variance of a continuous variable *X* with density function *f*(*x*)

$$Var(X) = E(X^2) - E(X)^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \left[\int_{-\infty}^{\infty} xf(x)dx\right]^2.$$

Example

- Consider the random variable X with density function

$$f(x) = \begin{cases} \dfrac{4+\sin(x)}{8\pi} & -\pi < x < \pi \\ 0 & \text{otherwise} \end{cases}$$



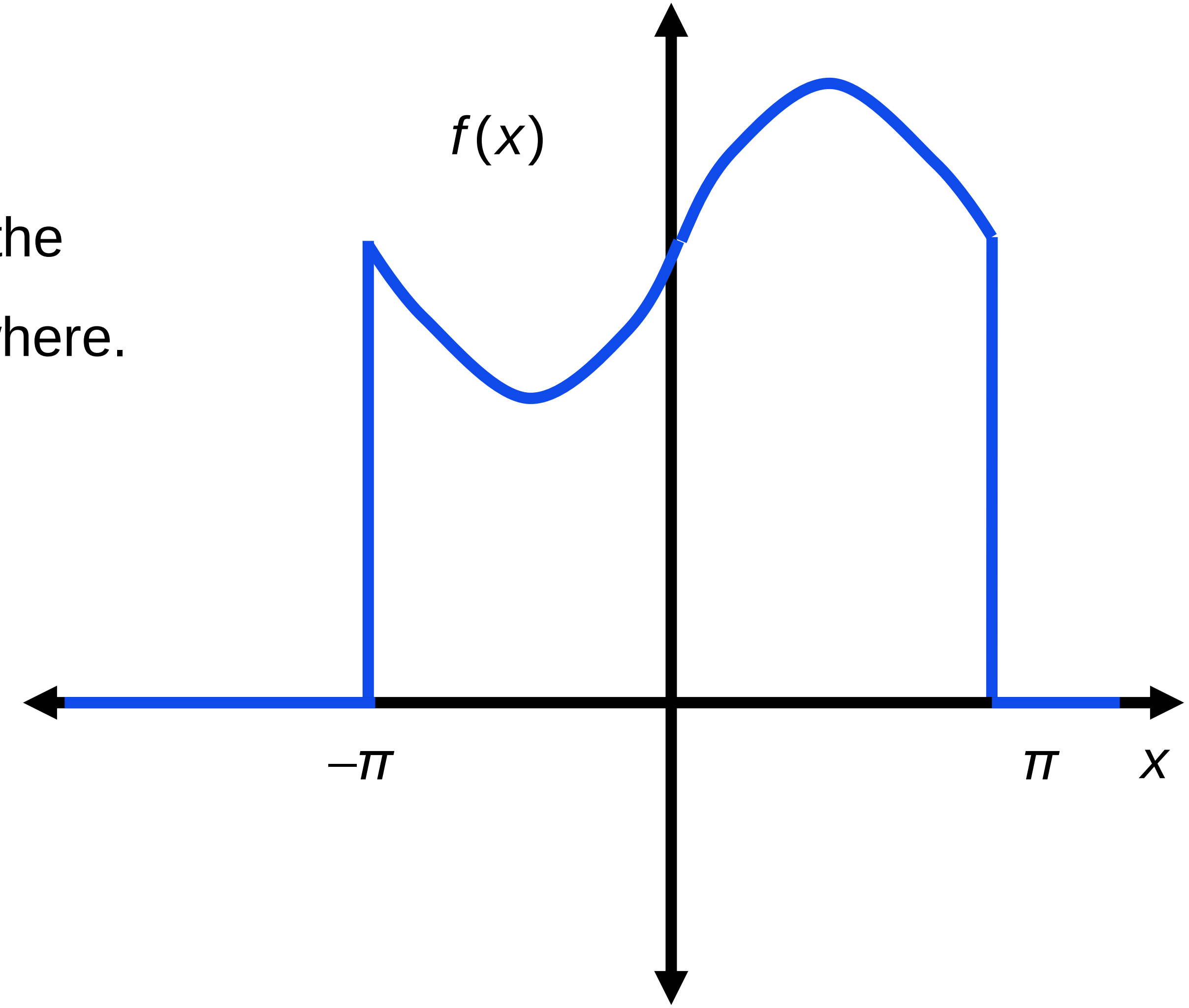
- We can verify that this is a valid density function

$$\text{since } \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\pi}^{\pi} \frac{4+\sin(x)}{8\pi}\,dx = \left[\frac{4x-\cos(x)}{8\pi}\right]_{-\pi}^{\pi} = 1$$

Example

- Visually, we can see that $E(X) > 0$ since from the graph we can observe that $f(x) \geq f(-x)$ everywhere.
- $E(X) = \int_{-\infty}^{\infty} xf(x)\, dx = \int_{-\pi}^{\pi} \frac{4x + x\sin(x)}{8\pi}\, dx$

$$= \left[\frac{2x^2 + \sin(x) - x\cos(x)}{8\pi} \right]_{-\pi}^{\pi} = \left[\frac{2\pi}{8\pi} \right] = \frac{1}{4}$$

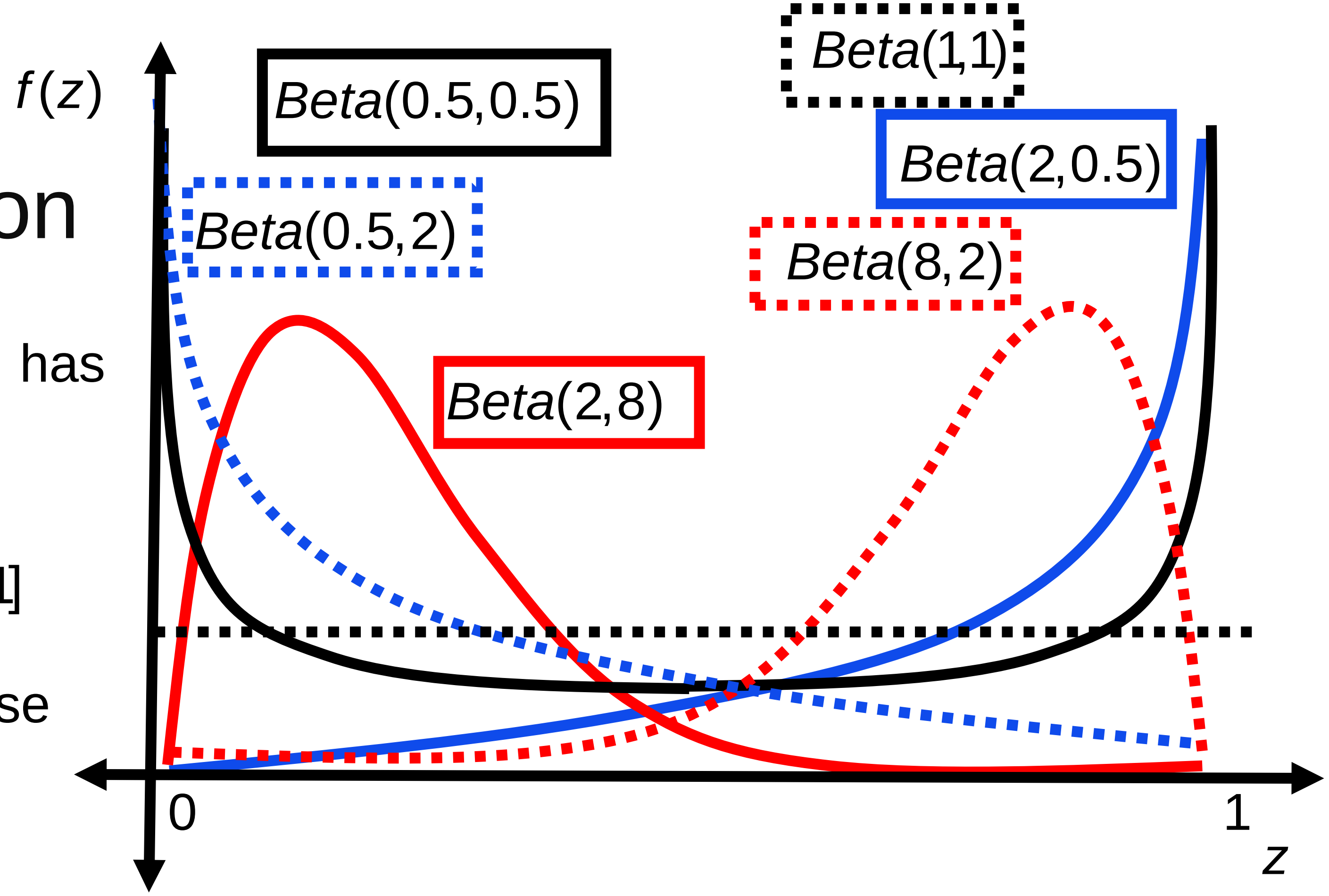


Example: Beta Distribution

- A **Beta** random variable $Z \sim Beta(\alpha, \beta)$ has probability density function

$$f(z) = \begin{cases} \dfrac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} z^{\alpha-1}(1-z)^{\beta-1} & z \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

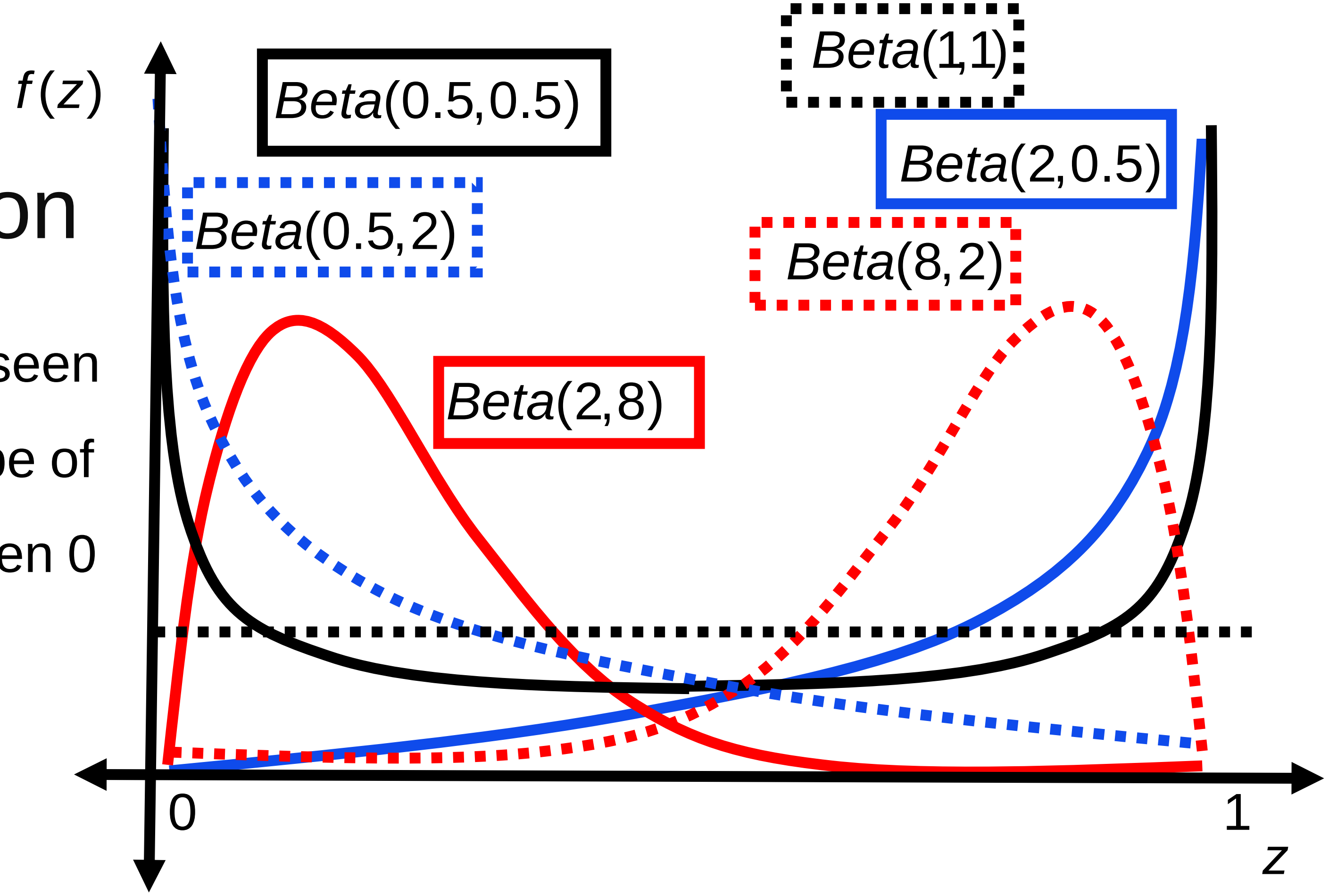
  where $\alpha > 0,\ \beta > 0.$



- Depending on the choice of parameters, it can describe very different distributions, some with mode (highest probability density) at 0, at 1 or at any value in between.

Example: Beta Distribution

- Beta random variables are commonly seen in Bayesian statistics (beyond the scope of 37161) as they only take values between 0 and 1 and so can be used to describe uncertainties in probability values.



Example: Beta Distribution

- In order to calculate $E(Z)=\int_{-\infty}^{\infty} zf(z)dz$, we can use the fact that

$$f(z)=\begin{cases} \dfrac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} z^{\alpha-1}(1-z)^{\beta-1} & z\in[0,1] \\ 0 & \text{otherwise} \end{cases}$$

is a valid probability density function.

- $\int_{-\infty}^{\infty} f(z)dz=1$ hence $\int_0^1 \dfrac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} z^{\alpha-1}(1-z)^{\beta-1}dz=1$ so $\int_0^1 z^{\alpha-1}(1-z)^{\beta-1}dz=\dfrac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$

Example: Beta Distribution

- $E(Z) = \int_{-\infty}^{\infty} z f(z) dz = \int_0^1 z \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} z^{\alpha-1}(1-z)^{\beta-1} dz = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} \int_0^1 z^{\alpha}(1-z)^{\beta-1} dz$
- We now use the fact that $\int_0^1 z^{\alpha-1}(1-z)^{\beta-1} dz = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$ to obtain $\int_0^1 z^{\alpha}(1-z)^{\beta-1} dz = \frac{(\alpha)!(\beta-1)!}{(\alpha+\beta)!}$
- Substituting this into gives $E(Z) = \int_{-\infty}^{\infty} z f(z) dz = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} \int_0^1 z^{\alpha}(1-z)^{\beta-1} dz$

$$= \left[\frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!}\right]\left[\frac{(\alpha)!(\beta-1)!}{(\alpha+\beta)!}\right] = \frac{\alpha}{\alpha+\beta}$$

Cumulative Density Functions

- For a continuous random variable with probability density function $f(x)$, the **cumulative density function** $F(x)$ is defined as $P(X \leq x)$.

- Just as a cumulative probability mass function for a discrete random variable is obtained by summing all probabilities up to and including a given value, the cumulative density function involves integrating all probability density up to and including a point.

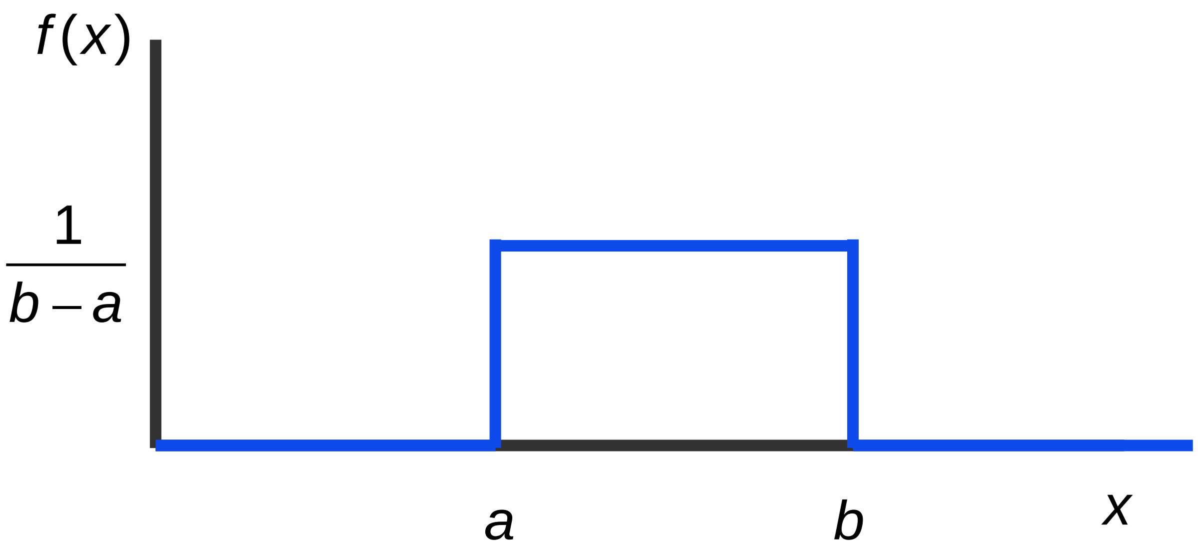
- This gives $P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt$.

- Similarly, if we know $F(x)$, we can easily obtain $f(x)$ since $F(x) = \int_{-\infty}^{x} f(t)dt$ implies that

$$f(x) = \frac{dF(x)}{dx}.$$

Example: Uniform Random Variable

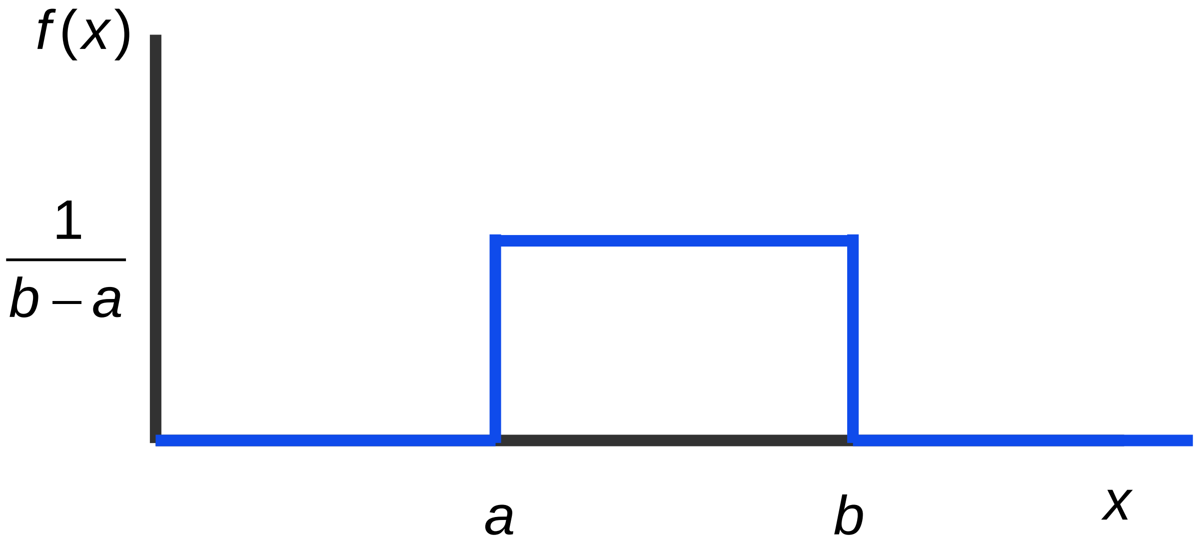
- Maybe the simplest type of continuous random variable is the **uniform** variable.
- For $a < b$ if $X \sim U[a,b]$ then X takes a value between a and b such that X is equally likely to be any two intervals of equal width.
- That is $f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$



Example: Uniform Random Variable

- We can see that this is a valid density function since

$$\int_{-\infty}^{\infty} f(x)dx = \int_a^b f(x)dx = \int_a^b \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_a^b = 1$$



- The expectation of this is

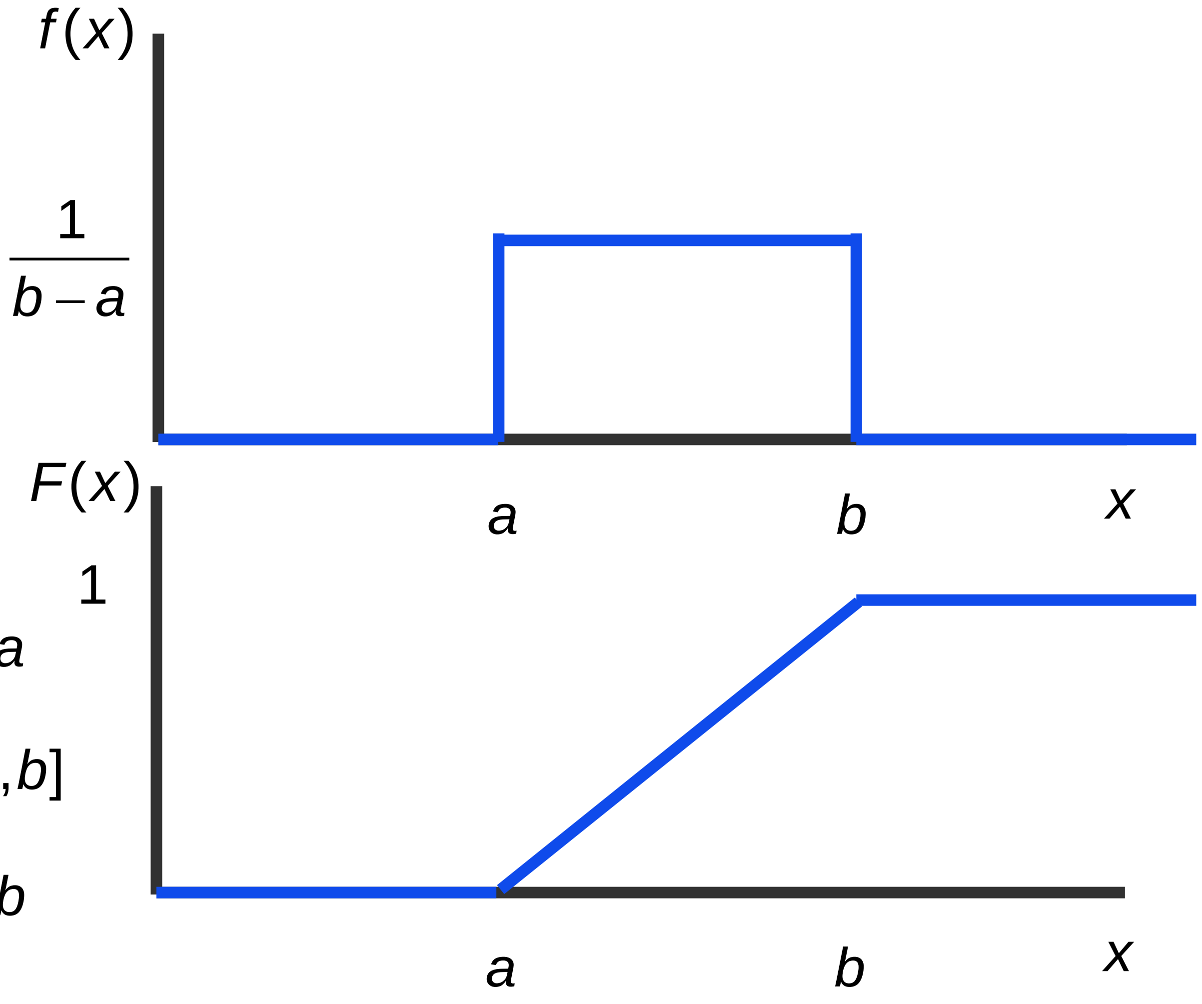
$$\int_{-\infty}^{\infty} xf(x)dx = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \left[\frac{(b^2-a^2)}{2(b-a)}\right] = \frac{b+a}{2}$$

Example: Uniform Random Variable

- The cumulative density function of is therefore

$$\int_{-\infty}^{x} f(t)dt = \begin{cases} 0 & x < a \\ \int_a^x \frac{1}{b-a}dt & x \in [a,b] \\ 1 & x > b \end{cases}$$

$$= \begin{cases} 0 & x < a \\ \left[\frac{t}{b-a}\right]_a^x & x \in [a,b] \\ 1 & x > b \end{cases} = \begin{cases} 0 & x < a \\ \left[\frac{x-a}{b-a}\right] & x \in [a,b] \\ 1 & x > b \end{cases}$$



Pareto Distribution

- A less commonly-seen variable is the **Pareto** variable.

- Named after the Italian economist Vilfredo Pareto, it is used to characterise strongly skewed data i.e. ones where small values are extremely likely and very large ones are extremely rare.

- Used in actuarial science and insurance modelling – for example, minor scrapes and car accidents are very common but low cost. Natural disasters (bushfires, earthquakes etc) are very rare but hugely costly.

- It also describes the size distribution of living organisms. For example, in the oceans, there are many billions of zooplankton for each fish and many billions of fish for each whale etc.

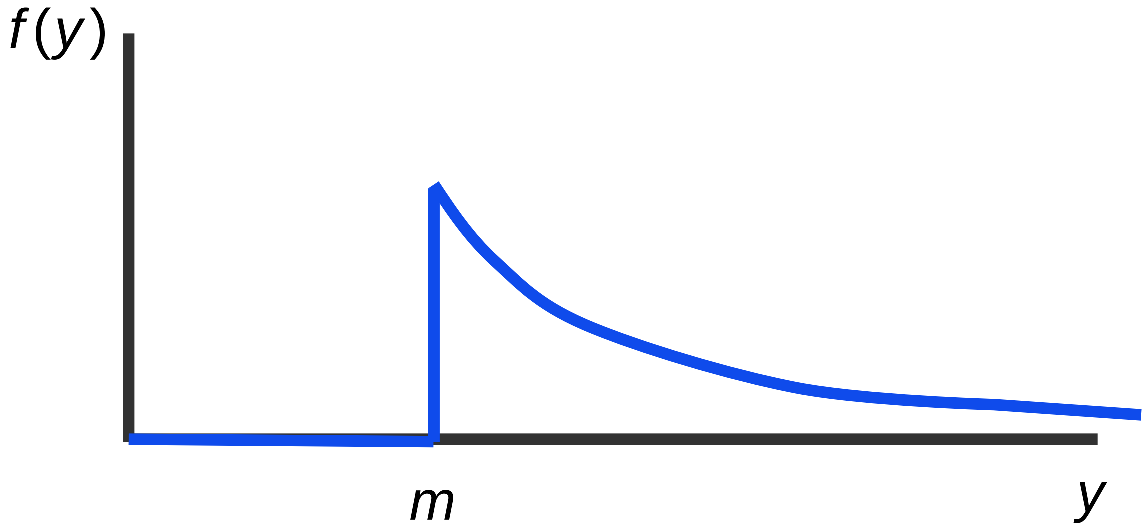
Vilfredo Pareto

(1848-1923)

Pareto Distribution

- If $Y \sim Pareto(m,\alpha)$ then (for $m > 0, \alpha > 0$)

$$f(y) = \begin{cases} \dfrac{\alpha m^{\alpha}}{y^{\alpha+1}} & y \in [m,\infty) \\ 0 & \text{otherwise} \end{cases}$$



- We can verify that this is a valid density function, since $\int_{-\infty}^{\infty} f(y)dy = \int_{m}^{\infty} f(y)dy = \int_{m}^{\infty} \dfrac{\alpha m^{\alpha}}{y^{\alpha+1}} dy = \left[-\dfrac{\alpha m^{\alpha}}{\alpha y^{\alpha}} \right]_{m}^{\infty} = 0 - \left(-\dfrac{\alpha m^{\alpha}}{\alpha m^{\alpha}} \right) = 1$

Pareto Distribution

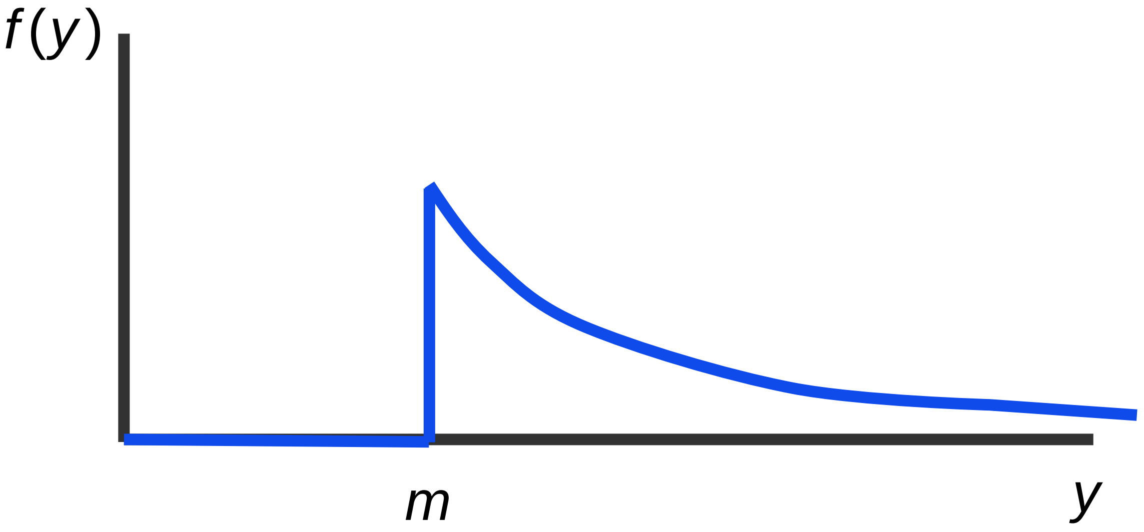
- The expectation of $Y \sim Pareto(m, \alpha)$ is therefore

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy = \int_{m}^{\infty} y f(y) dy$$

$$= \int_{m}^{\infty} y \frac{\alpha m^{\alpha}}{y^{\alpha+1}} dy = \int_{m}^{\infty} \frac{\alpha m^{\alpha}}{y^{\alpha}} dy$$

$$= \left[-\frac{\alpha m^{\alpha}}{(\alpha - 1) y^{\alpha - 1}} \right]_{m}^{\infty}$$

$$= 0 - \left(-\frac{\alpha m^{\alpha}}{(\alpha - 1) m^{\alpha - 1}} \right) = \frac{\alpha m}{\alpha - 1}. \quad \text{(provided } \alpha > 1\text{)}$$



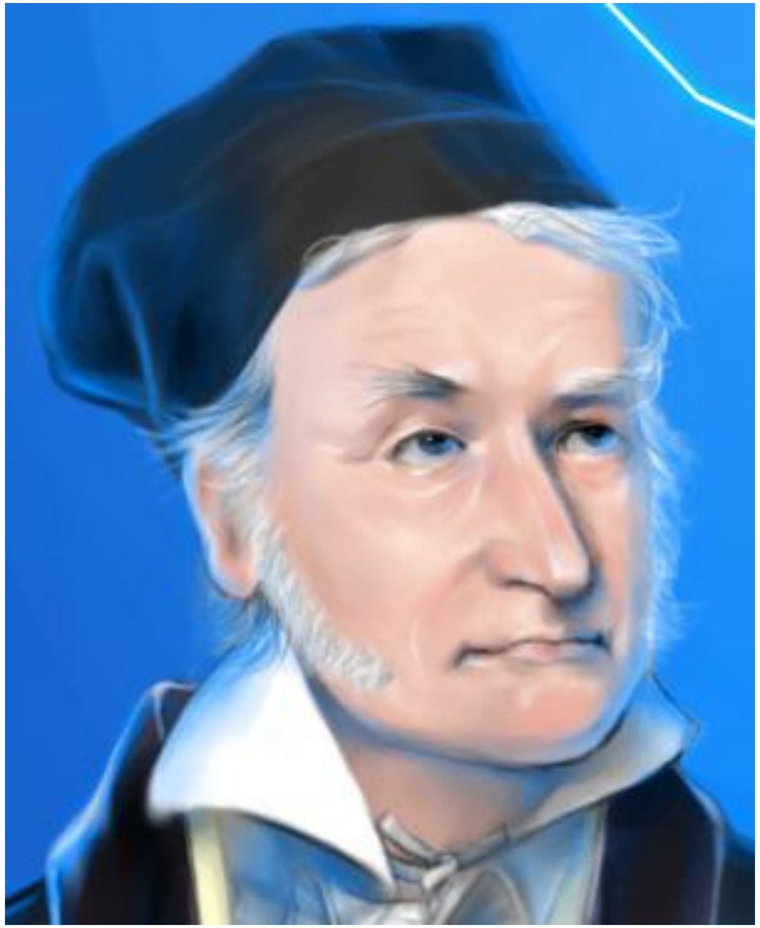
- Note, if $\alpha \leq 1$, then $E(Y) = \infty$.

Normal Distribution

- The most commonly seen variables used in statistics are **Normal** or **Gaussian** variables.

- This is the classic “bell curve” shape.

- It is beyond the scope of 37161, but the Central Limit Theorem tells us that (for most situations) if we average variables drawn from any other distribution with finite variance, then the averages themselves will be realisations of a Normal variable.



Karl Friedrich Gauss

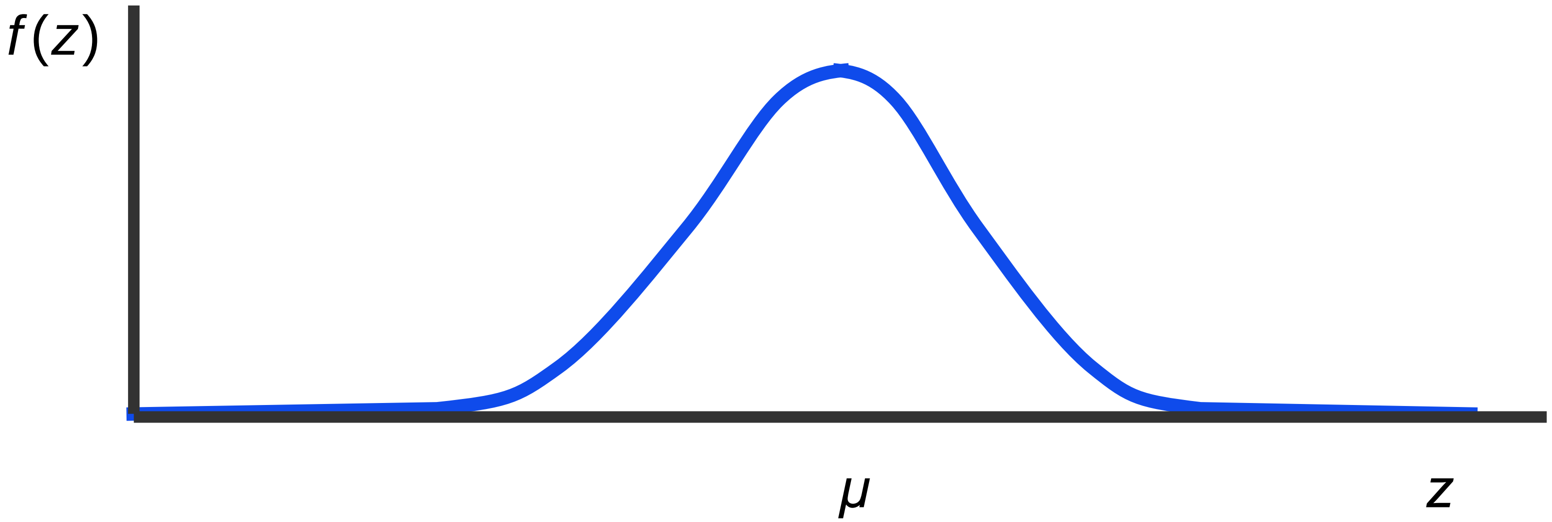
(*1777 – 1855*)

Normal Distribution

- If $Z \sim N(\mu, \sigma^2)$ then, for $\sigma^2 > 0$

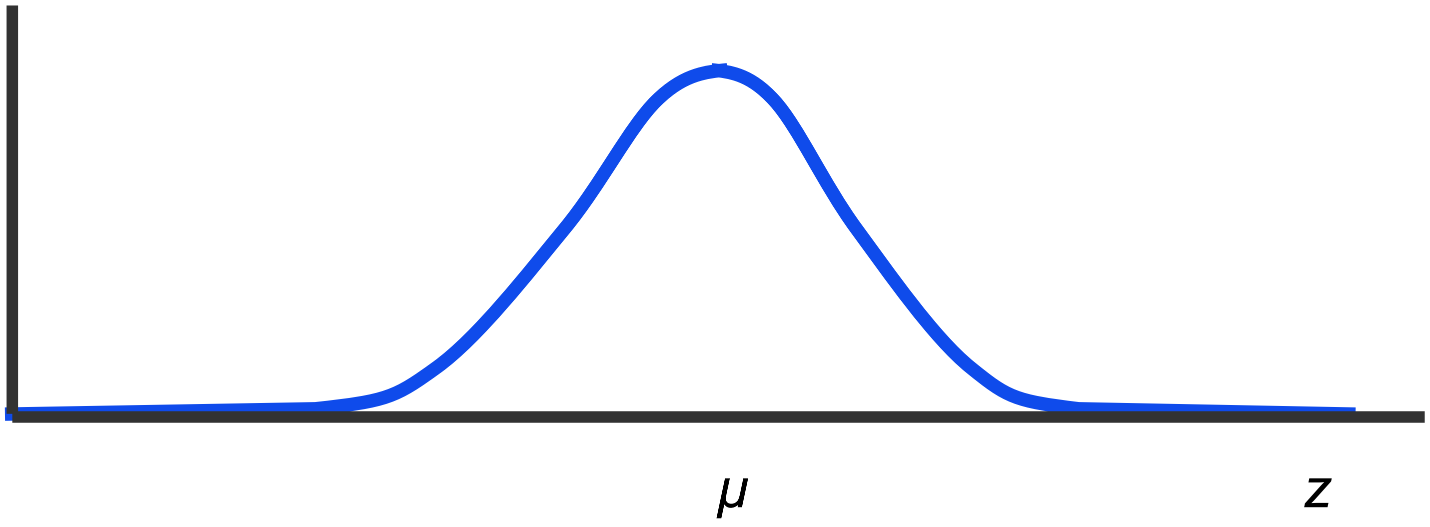
$$f(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(z-\mu)^2}{2\sigma^2}\right)}$$

where z can take any real value, positive or negative.



- This gives $E(Z) = \mu, Var(Z) = \sigma^2$

Normal Distribution



- We will not use this distribution a great deal in this subject, partly because there is no tidy closed form solution of $\int_a^b f(z)dz$ for almost all values of *a* and *b*. When working with the Normal distribution, numerical tables and/or computer packages are usually required.
- Even verification that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(z-\mu)^2}{2\sigma^2}\right)} dz = 1$ is a non-trivial task and requires transformation of variables into polar co-ordinates.