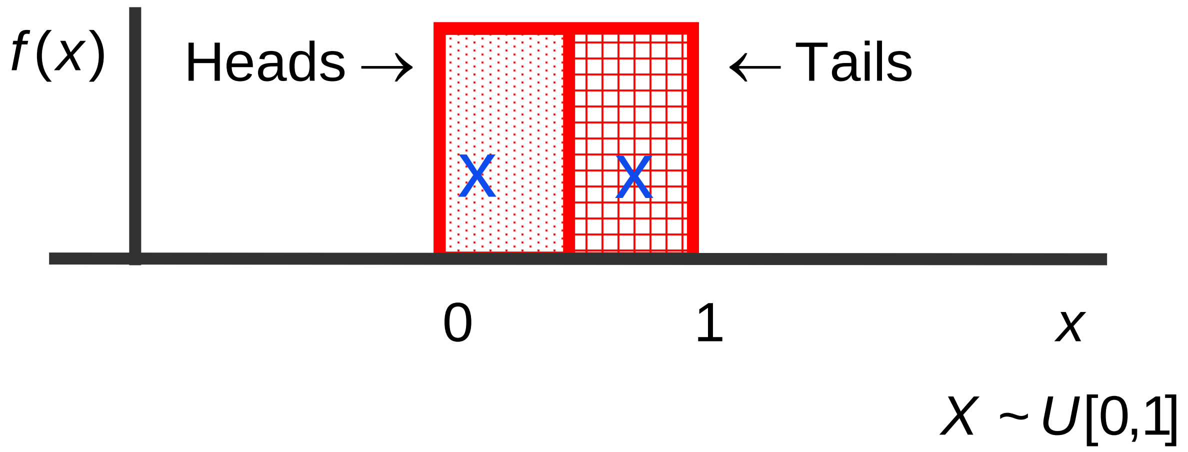
37262 Mathematical Statistics

Lecture 2

Simulating Discrete Events

- We saw in Lecture 1 how we could use realisations of a $U[0,1]$ variable to simulate realisations of flipping a fair coin.

- Using a realisation, u, of a $U[0,1]$ variable, if we call the outcome Heads if $u \le 0.5$ and Tails if $u > 0.5$, then we have a random experiment which returns Heads or Tails, each with probability 0.5.



Cumulative Probability Functions

- For a discrete random variable X with probability mass function $f(x) = P(X = x)$, the **cumulative probability function** is defined as $F(x) = P(X \leq x) = \sum_{k \leq x} P(X = k)$.
- In other words, for any input, the function gives the value that the random variable takes a value less than or equal to that input.
- For example, consider the random variable X which takes the value of the number shown when rolling one regular fair six-sided die.
- The probability mass function is clearly $P(X = k) = \begin{cases} 1/6 & k = 1 \\ 1/6 & k = 2 \\ 1/6 & k = 3 \\ 1/6 & k = 4 \\ 1/6 & k = 5 \\ 1/6 & k = 6 \\ 0 & \text{otherwise} \end{cases}$

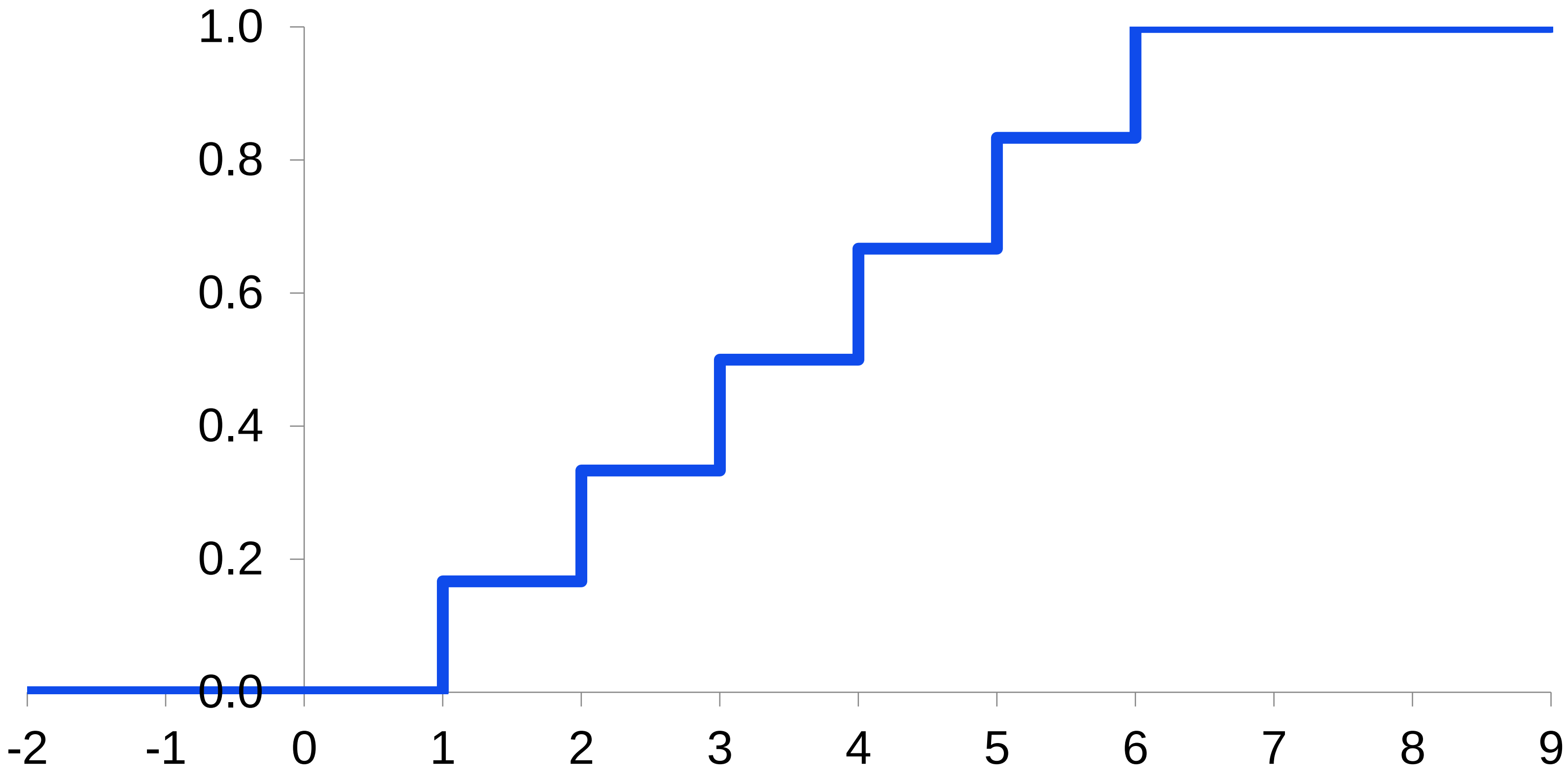
Cumulative Probability Functions

- When rolling a die, there is no chance of a result < 1.
- The chance of a result < 2 is 1/6.
- The chance of a result < 3 is 2/6 (or 1/3.)
- For $P(X=k)=\begin{cases} 1/6 & k=1 \\ 1/6 & k=2 \\ 1/6 & k=3 \\ 1/6 & k=4 \\ 1/6 & k=5 \\ 1/6 & k=6 \\ 0 & \text{otherwise} \end{cases}$, we have $P(X\leq k)=\begin{cases} 0 & k\in(-\infty,1) \\ 1/6 & k\in[1,2) \\ 1/3 & k\in[2,3) \\ 1/2 & k\in[3,4) \\ 2/3 & k\in[4,5) \\ 5/6 & k\in[5,6) \\ 1 & k\in[6,\infty) \end{cases}$

Cumulative Probability Functions

- Plotting up the cumulative probability function:

$$P(X \leq k) = \begin{cases} 0 & k \in (-\infty, 1) \\ 1/6 & k \in [1, 2) \\ 1/3 & k \in [2, 3) \\ 1/2 & k \in [3, 4) \\ 2/3 & k \in [4, 5) \\ 5/6 & k \in [5, 6) \\ 1 & k \in [6, \infty) \end{cases}$$



Cumulative Probability Functions

- We can generate realisations of a discrete random variable X from realisations of a $U[0,1]$ variable and the cumulative probability function .

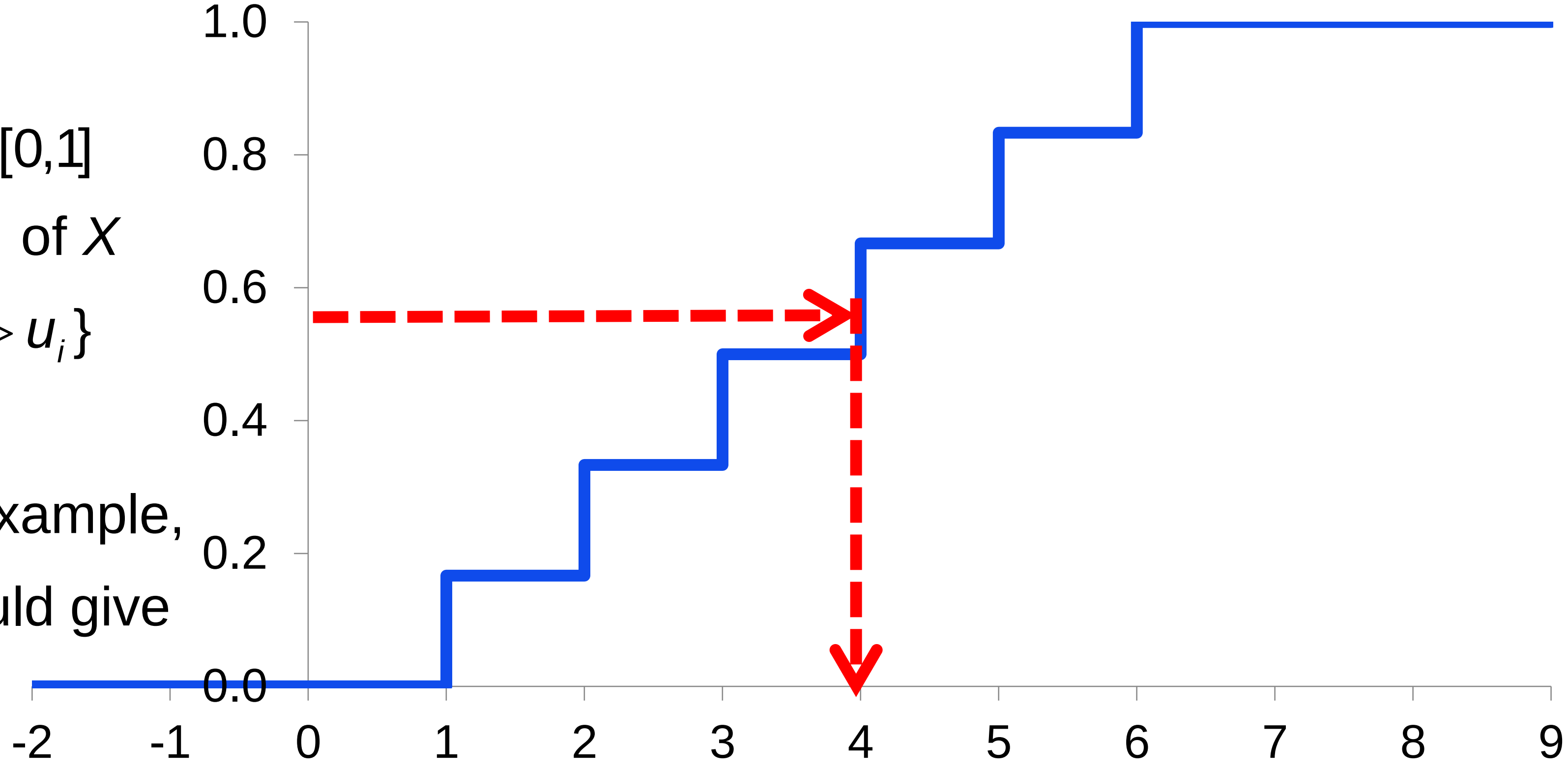
- Given a realisation , u_i of $U[0,1]$

  we generate a realisation x_i of X

  where $x_i = \min\{k : P(X \le k) > u_i\}$

- For example, with the die example,

  a realisation $u_i = 0.571$ would give $x_i = 4$.



Cumulative Probability Functions

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

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

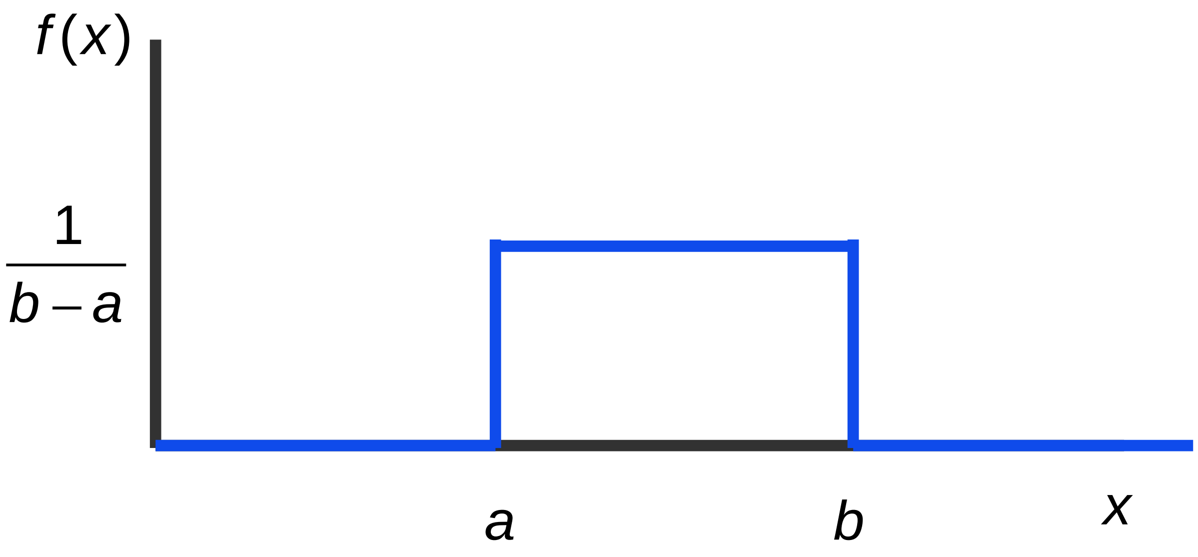
$$f(x) = \frac{dF(x)}{dx}.$$

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.

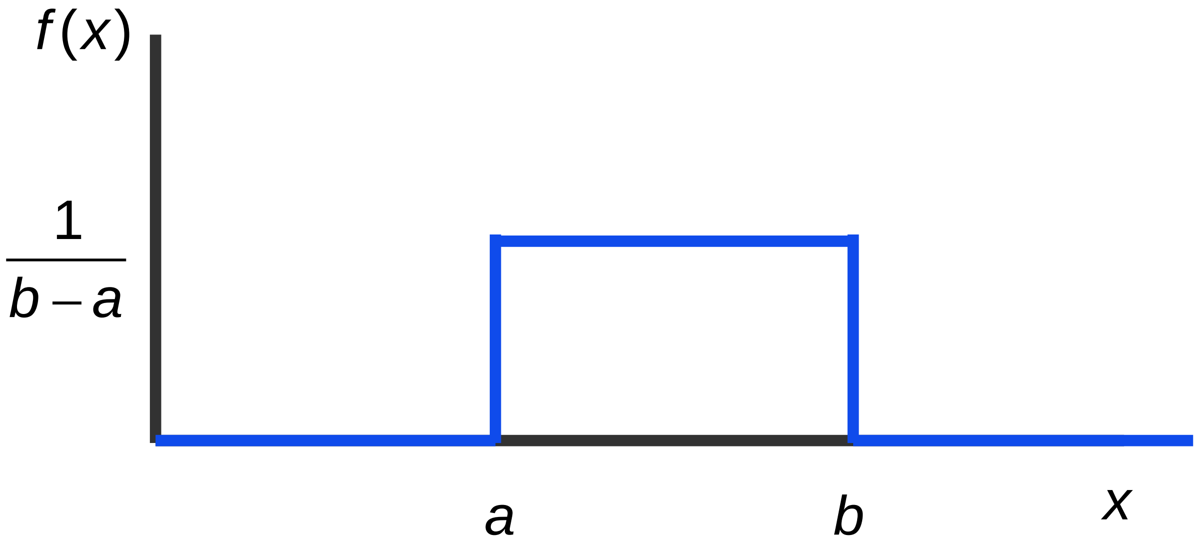
- That is $f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$



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

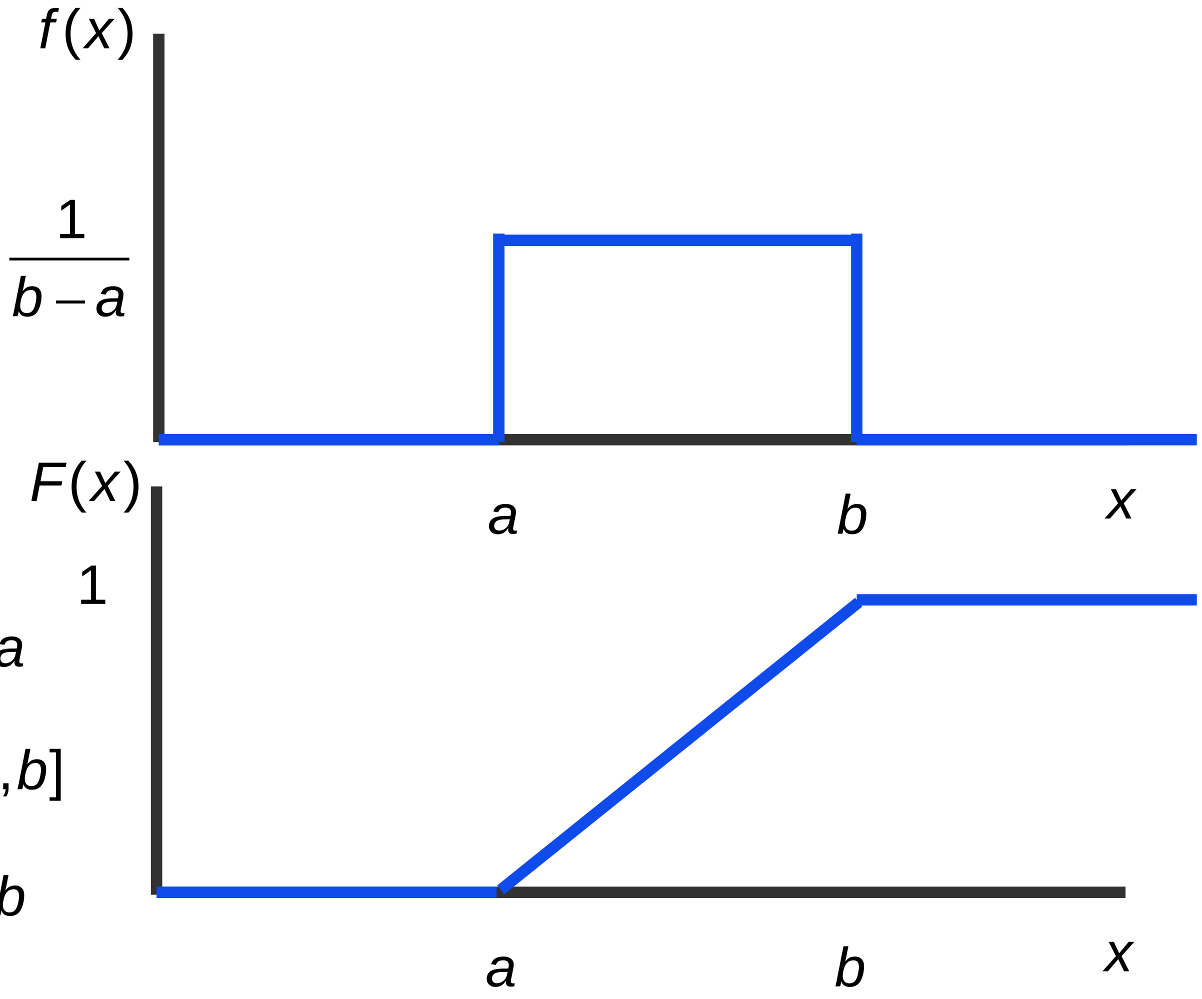
$$\int_{-\infty}^{\infty} xf(x)dx = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx = \left[\frac{x^2}{2(b-a)}\right]_a^b = \left[\frac{(b^2-a^2)}{2(b-a)}\right] = \frac{b+a}{2}$$

Example: Uniform Random Variable

- The cumulative probability function of X is therefore

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

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



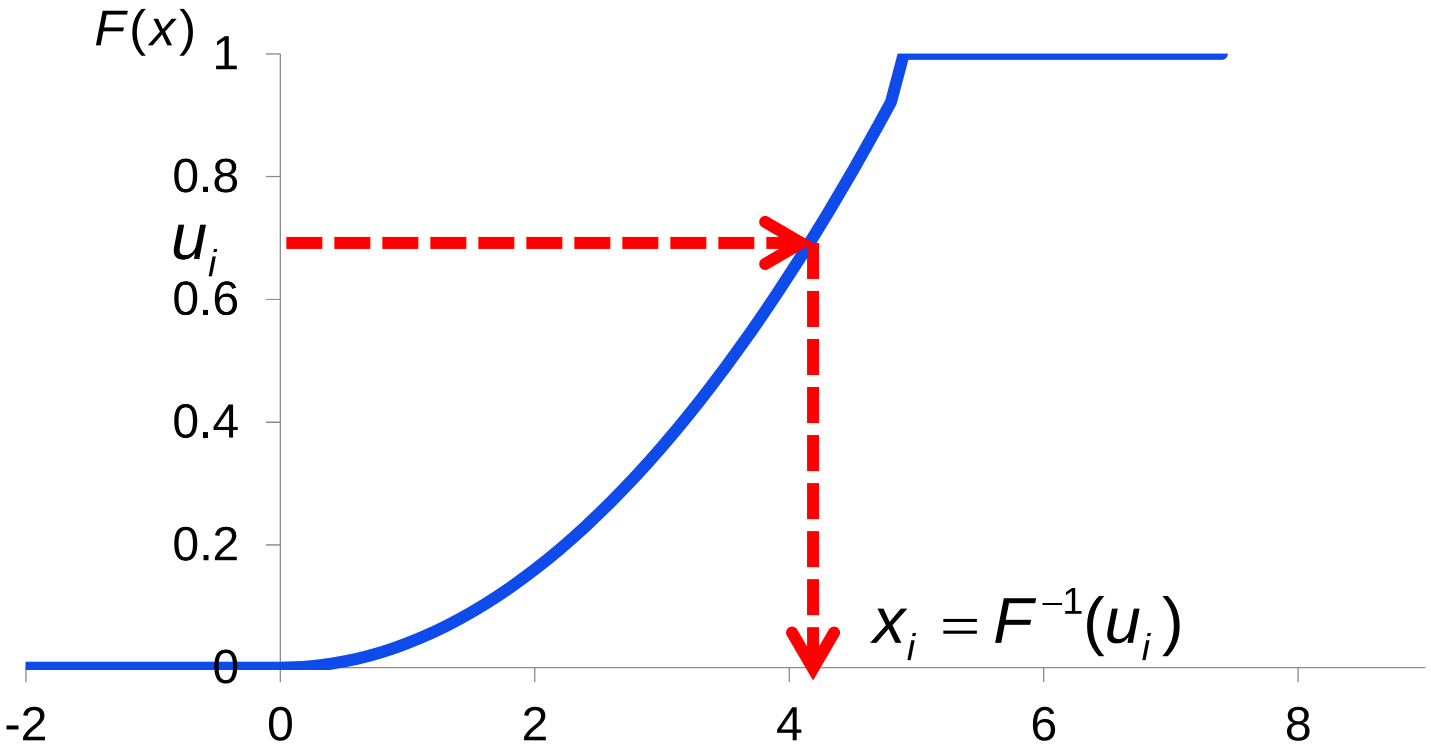
Simulating Continuous Random Variables

- We can generate realisations of a discrete random variable X from realisations of a $U[0,1]$ variable and the cumulative probability function

- Given a realisation u_i of $U[0,1]$, we generate a realisation x_i of X where $u_i = F(x_i)$.

- We do this via finding the inverse function for the cumulative function.

- We set $x_i = F^{-1}(u_i)$



Example: Exponential Random Variables

- A random variable $W \sim \exp(\lambda)$ is an **exponential** random variable if it has the density function

$$f(w) = \begin{cases} \lambda e^{-\lambda w} & w \in [0,\infty) \\ 0 & \text{otherwise} \end{cases}$$

- It is commonly used to model inter-event times, when events occur independently of each other in time.

- For example, if hospital admissions occur independently of each other, but with an average rate of λ per hour, then the time between two successive admissions (in hours) $\sim \exp(\lambda)$.

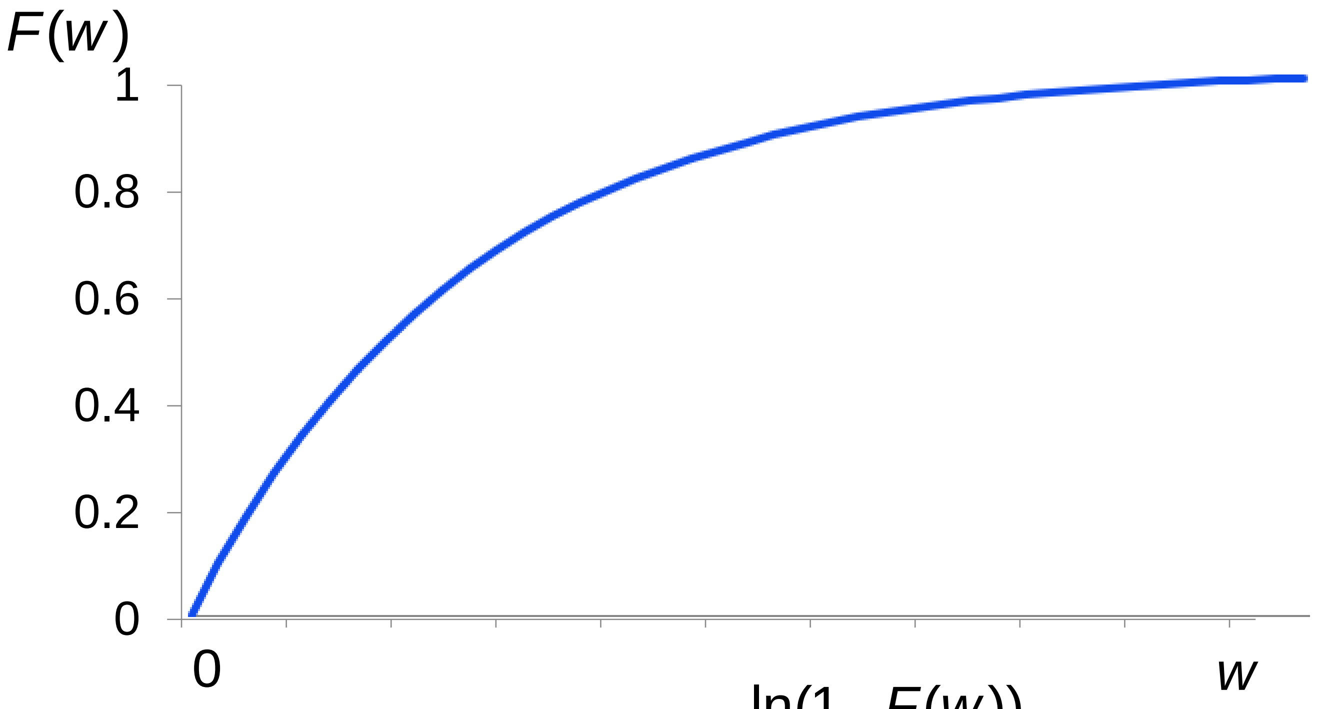
- Integration by parts gives the expectation of the variable as $E(W) = \int_0^\infty w\lambda e^{-\lambda w}\,dw =$

$$\int_0^\infty w\lambda e^{-\lambda w}\,dw = \left[-\frac{w^2}{2}e^{-\lambda w}\right]_0^\infty + \int_0^\infty e^{-\lambda w}\,dw \quad = 0 + \left[-\frac{e^{-\lambda w}}{\lambda}\right]_0^\infty = \frac{1}{\lambda}$$

Example: Exponential Random Variables

- The cumulative probability function of $W \sim \exp(\lambda)$ is therefore

$$\int_{-\infty}^{w} f(t)dt = \int_{0}^{w} \lambda e^{-\lambda t} dt \quad = \left[-e^{-\lambda t} \right]_0^w \quad = 1 - e^{-\lambda w}$$



- To generate realisations of $W \sim \exp(\lambda)$ from $U[0,1]$ realisations, we need to invert $F(w)$.

- $F(w) = 1 - e^{-\lambda w}$ hence $e^{-\lambda w} = 1 - F(w)$.

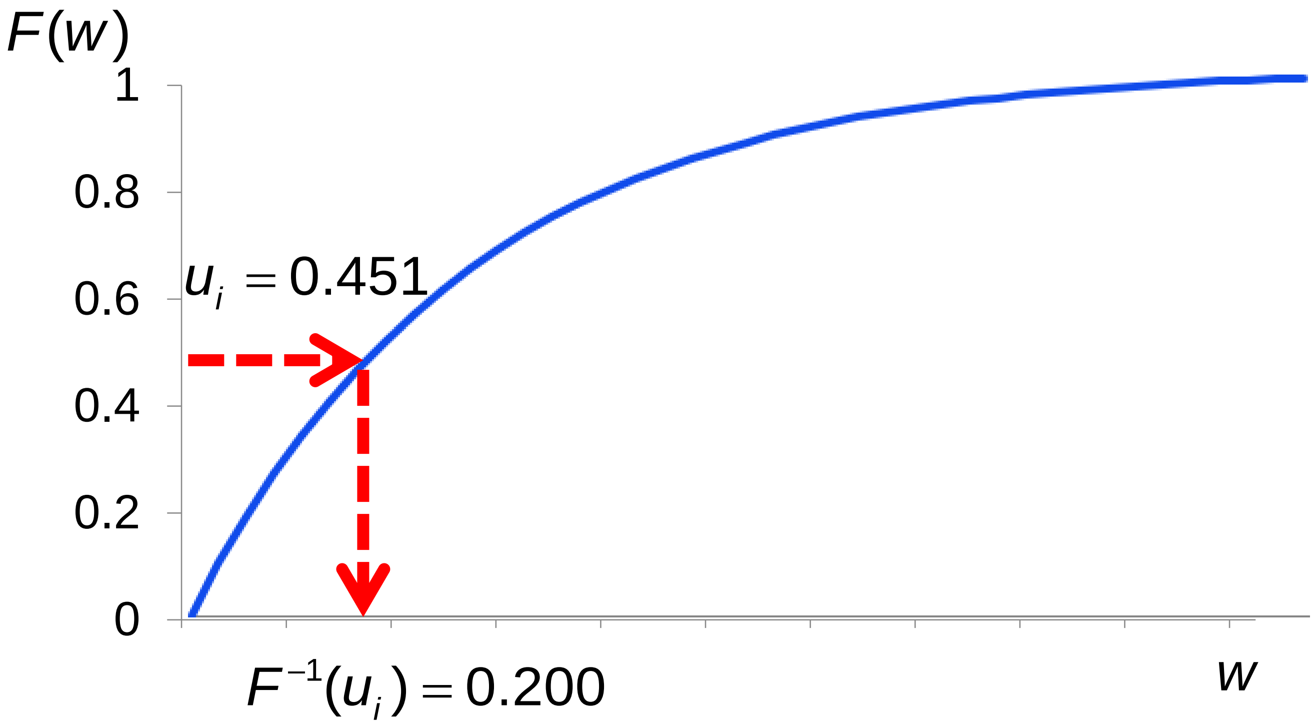
- Taking the logarithm of both sides gives $-\lambda w = \ln(1 - F(w))$ hence $w = -\dfrac{\ln(1-F(w))}{\lambda}$.

Example: Exponential Random Variables

- We have that $F^{-1}(w) = -\dfrac{\ln(1-w)}{\lambda}$.
- We therefore generate a realisation of W by from a realisation of $U[0,1]$, u,

  by $w = -\dfrac{\ln(1-u)}{\lambda}$
- For example, a realisation of $W \sim \exp(3)$ corresponding to $u_i = 0.451$ would be

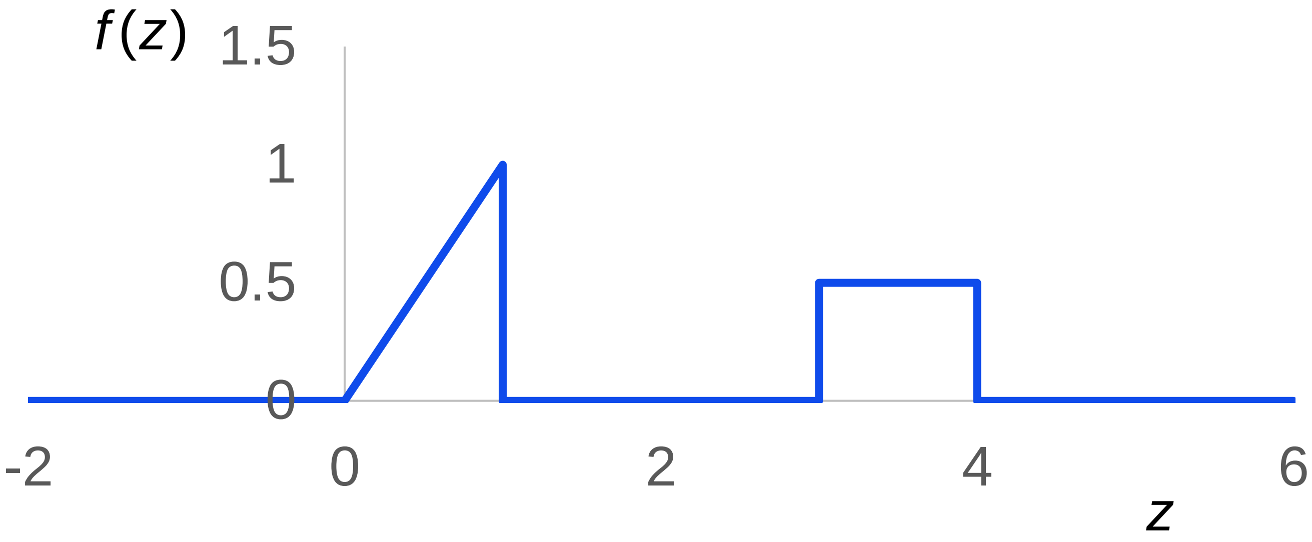
$$w_i = -\frac{\ln(1-0.451)}{3} \approx 0.200.$$



Example

- Let Z be a continuous random variable with probability density function

$$f(z) = \begin{cases} z & z \in [0,1] \\ 0.5 & z \in [3,4] \\ 0 & \text{otherwise} \end{cases}$$



- Clearly, there is no probability that Z is below 0 and it is certain that Z is no greater than 4 hence

$F(Z) = P(Z \leq z) = \begin{cases} 0 & z < 0 \\ 1 & z > 4 \end{cases}$. We integrate to find the other values.

Example

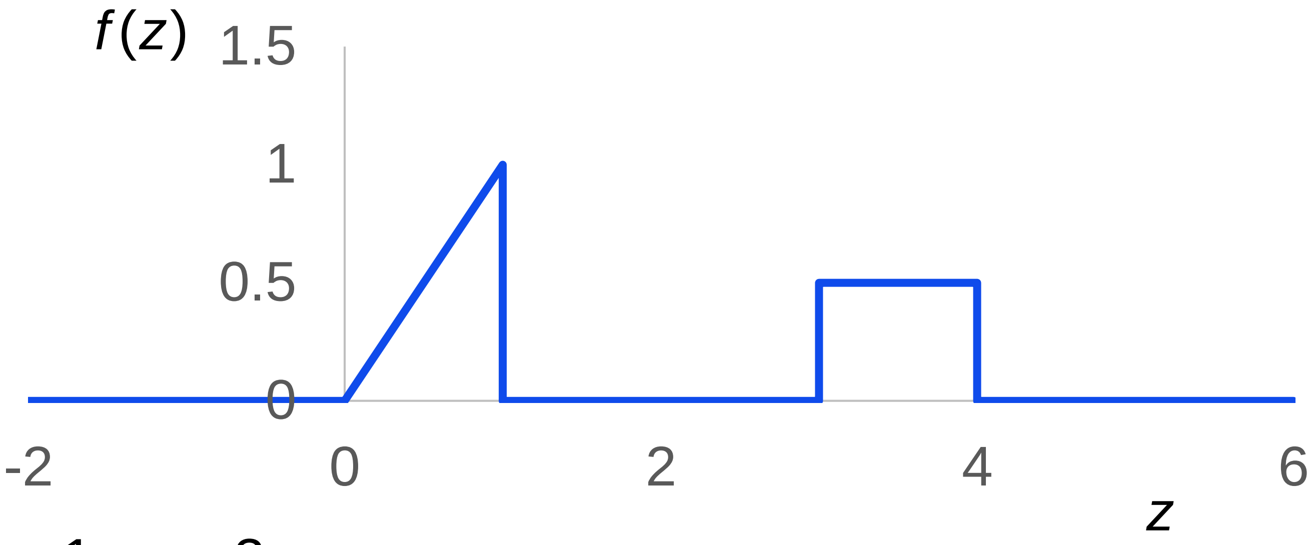
- For $z \in [0,1]$

$$F(z) = \int_0^z f(t)dt = \int_0^z t\,dt = \left[\frac{1}{2}t^2\right]_0^z = \frac{1}{2}z^2$$

- For $z \in [1,3]$, $F(z) = \left.\frac{1}{2}z^2\right|_{z=1} = \frac{1}{2}$

- For $z \in [3,4]$, $F(z) = \frac{1}{2} + \int_3^z f(t)dt$

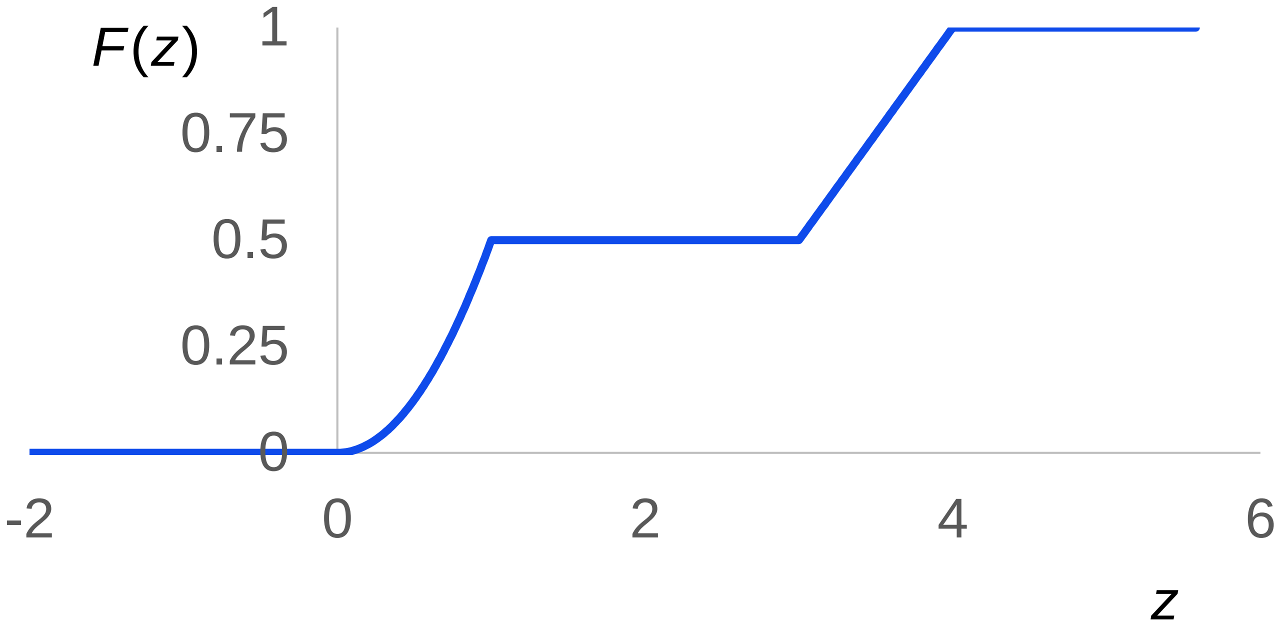
$$= \frac{1}{2} + \int_3^z \frac{1}{2}dt = \frac{1}{2} + \left[\frac{t}{2}\right]_3^z = \frac{1}{2} + \frac{z-3}{2}$$



Example

- Overall, this gives

$$F(z) = P(Z \le z) = \begin{cases} 0 & z < 0 \\ \frac{1}{2}z^2 & z \in [0,1] \\ \frac{1}{2} & z \in [1,3] \\ \frac{z-2}{2} & z \in [3,4] \\ 1 & z > 4 \end{cases}$$



Example

- Inverting $$F(z) = P(Z \le z) = \begin{cases} 0 & z < 0 \\ \frac{1}{2}z^2 & z \in [0,1] \\ \frac{1}{2} & z \in [1,3] \\ \frac{z-2}{2} & z \in [3,4] \\ 1 & z > 4 \end{cases}$$ we obtain $$F^{-1}(z) = \begin{cases} \sqrt{2z} & z \in \left[0, \frac{1}{2}\right] \\ 2z+2 & z \in \left[\frac{1}{2}, 1\right] \end{cases}$$

- We could therefore generate realisations of Z by simulating realisations $\{u_1, u_2, u_3, \ldots\}$ of $U[0,1]$

  and setting $$z_i = \begin{cases} \sqrt{2u_i} & u_i \in \left[0, \frac{1}{2}\right] \\ 2u_i + 2 & u_i \in \left[\frac{1}{2}, 1\right] \end{cases}$$

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.

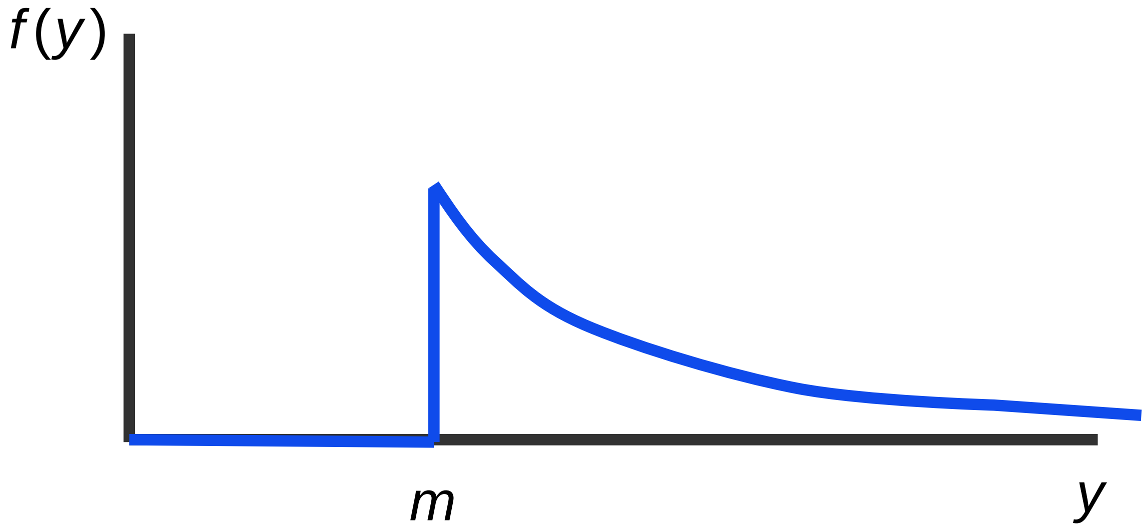
Vilfredo Pareto

(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 $\displaystyle\int_{-\infty}^{\infty} f(y)dy = \int_{m}^{\infty} f(y)dy = \int_{m}^{\infty} \frac{\alpha m^{\alpha}}{y^{\alpha+1}} dy = \left[-\frac{\alpha m^{\alpha}}{\alpha y^{\alpha}}\right]_{m}^{\infty} = 0 - \left(-\frac{\alpha m^{\alpha}}{\alpha m^{\alpha}}\right) = 1$

Example: Pareto Distribution

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

$$F(y)=\int_{-\infty}^{y} f(t)dt=\int_{m}^{y}\frac{\alpha m^{\alpha}}{t^{\alpha+1}}dt \quad =\alpha m^{\alpha}\int_{m}^{y}\frac{1}{t^{\alpha+1}}dt=\alpha m^{\alpha}\left[\frac{-1}{\alpha t^{\alpha}}\right]_{m}^{y} \quad =\alpha m^{\alpha}\left[\frac{1}{\alpha m^{\alpha}}-\frac{1}{\alpha y^{\alpha}}\right] \quad =\left[1-\frac{m^{\alpha}}{y^{\alpha}}\right]$$

- Inverting $F(y)=\left[1-\frac{m^{\alpha}}{y^{\alpha}}\right]$ gives $\frac{m^{\alpha}}{y^{\alpha}}=[1-F(y)]$ hence $\frac{m^{\alpha}}{[1-F(y)]}=y^{\alpha}$

- We therefore have that the inverse function is $F^{-1}(y)=\frac{m}{[1-y]^{\frac{1}{\alpha}}}$

- To generate realisations of a Pareto variable, we would simply evaluate $F^{-1}(u)=\frac{m}{[1-u]^{\frac{1}{\alpha}}}$ where u is drawn from a U[0,1] variable.

Acceptance-Rejection Methods

- We have already seen the convolution method, which allows us to simulate realisations of one random variable by summing other (easier to simulate) variables.
- We saw that a realisation of a binomial variable $Bin(n,p)$ could be obtained by summing n independent realisations of a Bernoulli variable, $Bern(p)$.
- Convolutions are used when summing a known number of variables.
- In some cases, we instead wish to see how many variables need to be summed before some criterion is reached.
- For example, counting how many independent realisations of a $Bern(p)$ variable until the first 1 is seen would give a realisation of a geometric variable $Geo(p)$.
- These are **acceptance-rejection** methods.

Poisson Distribution

- A Poisson variable $X \sim Poi(\lambda)$ is closely related to an exponential variable $W \sim \exp(\lambda)$.
- If events occur independently of each other, with expectation λ per unit time, then X gives the total number of events in unit time and W gives the time until the next event.
- We can therefore use simulated realisations of W to obtain realisations of X.
- We repeatedly simulate from $W \sim \exp(\lambda)$ variables and "accept" the next arrival only if it does not take the total time past 1. We stop this process when we first have to reject an arrival.
- This is considerably simpler than simulating directly from a Poisson probability mass function

$$P(N = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!} & k \in \{0,1,2,3,\ldots\} \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution

- For example, consider realisations of $X \sim \exp(4)$ which give

  $x_1 = 0.0198, x_2 = 0.4127, x_3 = 0.2557, x_4 = 0.1207, x_5 = 0.4655$

- This gives the first arrival after 0.0198, the second after 0.0198 + 0.4127 = 0.4325 and so on…

- The fourth arrival comes after $x_1 + x_2 + x_3 + x_4 = 0.8089$.

- The fifth comes after $0.8089 + x_5 = 0.8089 + 0.4655 = 1.2744$

- This implies only four events occur during unit time, hence our realisation of $X \sim Poi(4)$ is $x = 4$.