Theory of Inference - Notes

Dom Hutchinson

January 27, 2020

Contents

1 Motivation

Remark 1.1 - *General Idea*
Learn something about the world using data & statistical models.

Definition 1.1 - *Statistical Models*
Statistical Models describe the way in which data is generate. They depend upon *unknown* constant parameters, $\boldsymbol{\theta}$, and subsidiary information (known data & parameters).

Definition 1.2 - *Parameteric Statistical Inference*
Parameteric Statistical Inference is the process of taking some data & learning the *unknown* parameters of the model which generated it.
Definition 1.3 - *Parameteric Models*
A *Parameteric Model* is a statistical model whose pdf depends on some unknown parameter.
A *Semi-Parameteric Models* is a statistical models which contains unknown functions, as well as unknown parameters.
A *Non-Parameteric Model* has no parameters and thus makes minimal assumptions about how the data was generated.

Proposition 1.1 - *Inferential Questions*
When performing *Statistical Inference* we wish to answer the following questions

i) *Confidence Intervals & Credible Intervals* - What range of parameter valeus are consistent with the data?

ii) *Hypothesis Testing* - Are some pre-specified valeus (or restrictions) for the parameters consistent with the data?

iii) *Model Checking* - Could our model have generated the data at all?

iv) *Model Selection* - Which of several alternative odels could most plausibly have generated the data?

v) *Statistical Design* - How could we better arrange teh data gathering process to improve the answers to the preceding questions?

1.1 Examples

Example 1.1 - *Mean Annual Temperatures*
Consider a dataset of the mean annual temperature in New Haven, Conneticut.
Suppose we plot it in a histogram & notice that it fits a bell curve, then we may assume the data fits a simple model where each data point is observed independently from a $\mathcal{N}(\mu,\sigma^2)$ distribution with μ,σ^2 unknown.
Then the pdf for each data point, y_i, is

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

The pdf for the whole data set, $\mathbf{y}$, is the joint pdf of each data point since we assume iid

$$f(\mathbf{y})=\prod_{i=1}^N f(y_i)$$

Now suppose we notice that the histogram is *heavy tailed* relative to a normal distribution.
A better model might be

$$\frac{y_i-\mu}{\sigma}\sim t_\alpha$$

where μ, σ^2, α are unknown.
This means the pdf of the whole data set is

$$f(\mathbf{y}) = \prod_{i=1}^{N} \frac{1}{\sigma} f_{t_\alpha}\left(\frac{y_i - \mu}{\sigma}\right)$$

by *standard transformation theory.*

Example 1.2 - *Hourly Air Temperature*
Consider a dataset of the air temperature, a_i, measured at hourly intervals, t_i, over the course of a week.
The temperature is believed to follow a daily cycle, with a long-term dift over the course of the week and to be subject to random autocorrelated depatures from this overall pattern.
A suitable model might be

$$a_i = \underbrace{\theta_0 + \theta_1 t_i}_{\text{Long-Term Drift}} + \underbrace{\theta_2 \sin(2\pi t_i/24) + \theta_3 \cos(2\pi t_i/24)}_{\text{Daily Cycle}} + \underbrace{e_i}_{\text{Auto Correlation}}$$

where $e_{i+1} := \rho r_i + \varepsilon_i$ with $|\rho| < 1$ & $\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$.
This means $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ & $\mathbf{a} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ with $\Sigma_{i,j} = \frac{\rho^{|1-j|}\sigma^2}{1-\rho}$.
Thus, the pdf of the data set, $\mathbf{a}$, is

$$f(\mathbf{a}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(\mathbf{a}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{a}-\boldsymbol{\mu})}$$

Example 1.3 - *Bone Marrow*
Consider a dataset produced 23 patients suffering from non-Hodgkin's Lymphoma are split into two groups, each recieving a different treatment. We wish to test whether one of these treatments is more efficitive than the other.
For each patient the days between treatment & relapse was recorded. We have some *censored data* as the patient had not relapsed by the time of their last appointment.
Consider using an exponential distribution to model the times to relapse with parameters θ_a & θ_b respecitvely. We want to test if $\theta_a = \theta_b$.
We have the follow pdf for patients in group a

$$f_a(t_i) = \begin{cases} \theta_a e^{-\theta_a t_i} & \text{uncensored} \\ \int_{t_i}^{\infty} \theta_a e^{-\theta_a t_i} = e^{-\theta_a t_i} & \text{censored} \end{cases}$$

An equivalent pdf exists for patients in group b, with θ_b swapped in.
Thus the model for the whole data set, $\mathbf{t}$, is

$$f(\mathbf{t}) = \prod_{i=1}^{11} f_a(t_i) \prod_{i=12}^{23} f_b(t_i)$$

when patients $\{1, \dots, 11\}$ are in group a and the rest in group b.

2 Random Variables

Definition 2.1 - *Random Variable*
A *Random Variable* is a function from the sample space to the reals.

$$X : \Omega \to \mathbb{R}$$

Random Variables take a different value each time they are observed and thus we define distributions for the probability of them taking particular values.
Random Variables form the basis of models.

Definition 2.2 - *Cummulative Distribution*
The *Cummulative Distribution* function of a *Random Variable*, X, is the function $F_X(\cdot)$ st

$$\begin{array}{rcl} F_X(\cdot) & : & \mathbb{R} \to [0,1] \\ F_X(x) & := & \mathbb{P}(X \leq x) \quad = \quad \displaystyle\sum_{i=-\infty}^{x} \mathbb{P}(X = i) \\ & & \qquad\qquad\quad\; = \quad \displaystyle\int_{-\infty}^{x} f_X(x)dx \end{array}$$

The *Cummulative Distribution* is a monotonic function.

Remark 2.1 - *Continuous Cummulative Distribution*
If a *Cummulative Distribution* is *continuous* then $F_X(X) \sim \text{Uniform}[0,1]$.

Proof 2.1 - *Remark 2.1*

$$\begin{array}{rcl} F(X) & = & \mathbb{P}(X \leq x) \\ & = & \mathbb{P}(F(X) \leq F(x)) \\ \implies \mathbb{P}(F(X) \leq u) & = & u \text{ if } F \text{ is continuous} \end{array}$$

Definition 2.3 - *Quantile Function*
The *Quantile Function* of a *Random Variable* is the inverse function of the *Cumulative Distribution.*

$$\begin{array}{rcl} F_X^-(\cdot) & : & [0,1] \to \mathbb{R} \\ F_X^-(u) & := & \min\{x : F(x) \geq u\} \end{array}$$

If a distribution has a computable *Quantile Function* then we are able to generate random variable values by sampling from a uniform distribution & then passing that value into the *Quantile Function.*

Definition 2.4 - *(Q-Q) Plot*
Consider a data set $\{x_1, \dots, x_n\}$.
A *(Q-Q) Plot* of this data set plots the ordered data set, $\{x_{(1)}, \dots, x_{(n)}\}$, against the theoretical quantiles $F^-\left(\frac{i-.5}{n}\right)$.
The close this line is to $y = x$ the more likely it is the data was generated by this *Cummulative Distribtion.*
N.B. AKA *Quantile-Quantile Plot*
Definition 2.5 - *Probabiltiy Mass Function*
A *Probability Mass Function* returns the probabiltiy of a <u>discrete</u> random variable taking a particular value.

$$\begin{array}{rcl} f_X(\cdot) & : & \mathbb{R} \to [0,1] \\ f_X(x) & := & \mathbb{P}(X = x) \end{array}$$

Definition 2.6 - *Probability Density Function*
Since the probabiltiy of a *Continuous Random Variable* taking a specific value is zero we cannot use the *Probability Mass Function.*

$$\begin{array}{rcl} f_X(\cdot) & : & \mathbb{R} \to [0,1] \\ \mathbb{P}(a \leq X \leq b) & = & \displaystyle\int_a^b f(x)dx \end{array}$$

N.B. $F'_X(x) = f(x)$ when $F'_X(\cdot)$ exists.

Definition 2.7 - *Joint Probabiltiy Density Function*
Let X &Y be *Random Variables*.
The *Joint Probabiltiy Density Function* of X and Y is the function $f_{X,Y}(x,y)$ st

$$\mathbb{P}((X,Y)\in\Omega) = \iint_\Omega f_{X,Y}(x,y)dxdy$$

N.B. This can be seen as evaluation Ω in the $X-Y$ plane.

Definition 2.8 - *Marginal Distribution*
Let X & Y be *Random Variables* with *Joint Probability Density* $f_{X,Y}(\cdot,\cdot)$.
We can find the *Marginal Distribution* of X by evaluating the $f_{X,Y}$ at each value wrt Y.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Definition 2.9 - *Expected Value,* $\mathbb{E}$
The *Expected Value* of a *Random Variable*, X, is its mean value.

$$\begin{array}{rcll}
\mathbb{E}(X) & := & \displaystyle\int_{-\infty}^{\infty} xf(x)dx & \text{[Continuous]} \\
\mathbb{E}(g(X)) & := & \displaystyle\int_{-\infty}^{\infty} g(x)f(x)dx & \\
\\
\mathbb{E}(X) & := & \displaystyle\sum_{-\infty}^{\infty} xf(x) & \text{[Discrete]} \\
\mathbb{E}(g(X)) & := & \displaystyle\sum_{-\infty}^{\infty} g(x)f(x) &
\end{array}$$

Remark 2.2 - *Linear Transformations of Expected Value*

$$\mathbb{E}(a+bX) = a + b\mathbb{E}(X) \text{ where } a,b\in\mathbb{R}$$

Remark 2.3 - *Expected Value of Composed Random Variables*
Let X & Y be *Random Variables*. Then

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

If X & Y are *independent*. Then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Proof 2.2 - *Remark 2.3*

$$\begin{array}{rcl}
\mathbb{E}(X+Y) & = & \int(x+y)f_{X,Y}(x,y)dxdy \\
& = & \int xf_{X,Y}(x,y)dxdy + \int yf_{X,Y}(x,y)dxdy \\
& = & \mathbb{E}(X)+\mathbb{E}(Y) \\
\mathbb{E}(XY) & = & \int xyf_{X,Y}(x,y)dxdy \\
& = & \int xf_X(x)yf_Y(y)dxdy \text{ by independence} \\
& = & \int xf_X(x)dx\int yf_Y(y)dy \\
& = & \mathbb{E}(X)\mathbb{E}(Y)
\end{array}$$

Definition 2.10 - *Variance, σ^2*
The *Variance* of a *Random Variable*, X, is a measure of its spread around its expected value.

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

Remark 2.4 - *Linear Transformations of Variance*

$$\text{Var}(a + bX) = b^2\text{Var}(X) \text{ where } a, b \in \mathbb{R}$$

Proof 2.3 - *Remark 2.4*

$$\begin{array}{rcl} \text{Var}(a+bX) & = & \mathbb{E}[((a+bX) - (a - b\mu))^2] \\ & = & \mathbb{E}[b^2(X-\mu)^2] \\ & = & b^2\mathbb{E}[(X-\mu)^2] \\ & = & b^2\text{Var}(X) \end{array}$$

Definition 2.11 - *Co-Variance*
Co-Variance is a measure of the joint variability of two *Random Variables.*

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

N.B. If X & Y are independent then $\text{Cov}(X, Y) = 0$ since $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.
N.B. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.

Definition 2.12 - *Co-Variance Matrix, Σ*
Let $\mathbf{X} := \{X_1, \dots, X_n\}$ be a set of random variables.
A *Co-Variance Matrix* describes the *Variance* & *Co-Variance* of each combination of *Random Variables* in $\mathbf{X}$.

$$\Sigma := \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

N.B. $\Sigma_{ii} = \text{Var}(X_i)$ & $\Sigma_{ij} = \text{Cov}(X_i, X_j)$ for $i \neq j$. Σ is symmetric.

Remark 2.5 - *Linear Transformation of Covariance*

$$\Sigma_{AX+b} = A\Sigma A^T$$

Proof 2.4 - *Remark 2.5*

$$\begin{array}{rcl} \Sigma_{AX+b} & = & \mathbb{E}[(AX + \mathbf{b} - A\boldsymbol{\mu} - \mathbf{b})(AX + \mathbf{b} - A\boldsymbol{\mu} - \mathbf{b})^T] \\ & = & \mathbb{E}[(AX - A\boldsymbol{\mu})(AX - A\boldsymbol{\mu})^T] \\ & = & A\mathbb{E}[(X - \boldsymbol{\mu})(X - \boldsymbol{\mu})^T]A^T \\ & = & A\Sigma A^T \end{array}$$

Definition 2.13 - *Conditional Distribution*
Let X & Y be *Random Variables* with *Joint Probability Density* $f_{X,Y}(\cdot, \cdot)$.
Suppose we know that Y takes the value y_0 & we wish to establish the probability of X taking the value x.

$$f(X = x | Y = y_0) = \frac{f_{X,Y}(x, y_0)}{f_Y(y_0)}$$

assuming $f(y_0) > 0$.

Proof 2.5 - *Conditional Distribution*
We expect $f(X=x|Y=y_0) = kf_{X,Y}(x,y_0)$ for some constant k.
We know that for $kf_{X,Y}(x,y_0)$ to be a valid distribution it must integrate to one.

$$\begin{array}{rrcl} & k\int_{-\infty}^{\infty} f_{X,Y}(x,y_0)dx & = & 1 \\ \implies & kf_Y(y_0) & = & 1 \\ \implies & k & = & \frac{1}{f_Y(y_0)} \\ \implies & f(X=x|Y=y_0) & = & \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)} \end{array}$$

Proposition 2.1 - *Conditional Distributions with Three Random Variables*

$$\begin{array}{rcl} f(x,z|y) & = & f(x|z,y)f(z|y) \\ f(x,y,z) & = & f(x|y,z)f(z|y)f(y) \\ & = & f(x|y,z)f(y,z) \end{array}$$

Definition 2.14 - *Independent Random Variables*
Let X & Y be random variables.
X & Y are said to be *Statistically Independent* if the *Conditional Distribution* $f(x|y)$ is independent of y.
Thus

$$\begin{array}{rcl} f(x) = \int_{-\infty}^{\infty} f(x,y)dy & & \\ & = & \int_{-\infty}^{\infty} f(x|y)f(y)dy \\ & = & f(x|y)\int_{-\infty}^{\infty} f(y)dy \\ & = & f(x|y) \\ \implies f(x,y) & = & f(x|y)f_Y(y) = f_X(x)f_Y(y) \end{array}$$

Theorem 2.1 - *Bayes' Theorem*
Let X & Y be *Random Variables.*
Bayes' Theorem states that

$$f(X|Y) = \frac{f(Y|X)x(X)}{f(Y)}$$

Definition 2.15 - *First Order Markov Property*
Let $\mathbf{X} := \{X_1, \dots, X_n\}$ be a set of *Random Variables.*
The set $\mathbf{X}$ is said to have the *First Order Markov Property* if

$$f(X_i|\mathbf{X}_{\neg i}) = f(X_i|X_{i-1}) \text{ where } \mathbf{X}_{\neg i} := \mathbf{X}/\{X_i\}$$

Thus we can infer the *marginal distribution*

$$f(\mathbf{X}) = f(X_1)\prod_{i=2}^{N} f(X_i|X_{i-1})$$

0 Reference

0.1 Definitions

Definition 0.1 - *Heavy Tailed*

Definition 0.2 - *Censored Data*

0.2 Probability Distributions

Definition 0.3 - *β-Distribution*
Let $X \sim \text{Beta}(\alpha, \beta)$.
A *continuous* random variable with shape parameters $\alpha, \beta > 0$. Then

$$\begin{array}{rcl} f_X(x) & \propto & x^{\alpha-1}(1-x)^{\beta-1}\mathbb{1}\{x \in [0,1]\} \\ \mathbb{E}(X) & = & \dfrac{\alpha}{\alpha+\beta} \\ \text{Var}(X) & = & \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ \mathcal{M}_X(t) & = & 1 + \displaystyle\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha+r}{\alpha+\beta+r}\right)\frac{t^k}{k!} \end{array}$$

Definition 0.4 - *Bernoulli Distribution*
Let $X \sim \text{Bernoulli}(p)$.
A *discrete* random variable which takes 1 with probability p & 0 with probability $(1-p)$. Then

$$\begin{array}{rcl} p_X(k) & = & \begin{cases} 1-p & \text{if } k=0 \\ p & \text{if } k=1 \\ 0 & \text{otherwise} \end{cases} \\ P_X(k) & = & \begin{cases} 0 & \text{if } k<0 \\ 1-p & \text{if } k \in [0,1) \\ 1 & \text{otherwise} \end{cases} \\ \mathbb{E}(X) & = & p \\ \text{Var}(X) & = & p(1-p) \\ \mathcal{M}_X(t) & = & (1-p)+pe^t \end{array}$$

N.B. Often we define $q := 1-p$ for simplicity.

Definition 0.5 - *Binomial Distribution*
Let $X \sim \text{Binomial}(n, p)$.
A *discrete* random variable modelled by a *Binomial Distribution* on n independent events and rate of success p.

$$\begin{array}{rcl} p_X(k) & = & \binom{n}{k}p^k(1-p)^{n-k} \\ P_X(k) & = & \sum_{i=1}^{k}\binom{n}{i}p^i(1-p)^{n-i} \\ \mathbb{E}(X) & = & np \\ \text{Var}(X) & = & np(1-p) \\ \mathcal{M}_X(t) & = & [(1-p)+pe^t]^n \end{array}$$

N.B. If $Y := \sum_{i=1}^{n} X_i$ where $\mathbf{X} \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ then $Y \sim \text{Binomial}(n, p)$.

Definition 0.6 - *Categorical Distribution*
Let $X \sim \text{Categorical}(\mathbf{p})$.
A *discrete* random varaible where probability vector $\mathbf{p}$ for a set of events $\{1, \dots, m\}$.

$$f_X(i) = p_i$$

Definition 0.7 - *χ^2 Distribution*
Let $X \sim \chi^2_r$.

A *continuous* random variable modelled by the χ^2 *Distribution* with r degrees of freedom. Then

$$\begin{array}{rcl} f_X(x) &=& \frac{1}{2^{r/2}\Gamma(r/2)}x^{\frac{r}{2}-1}e^{-\frac{x}{2}} \\ F_X(x) &=& \frac{1}{\Gamma(k/2)}\gamma\left(\frac{r}{2},\frac{x}{2}\right) \\ \mathbb{E}(X) &=& r \\ \text{Var}(X) &=& 2r \\ \mathcal{M}_X(t) &=& \mathbb{1}\{t<\frac{1}{2}\}(1-2t)^{-\frac{r}{2}} \end{array}$$

N.B. If $Y := \sum_{i=1}^k Z_i^2$ with $\mathbf{Z} \overset{\text{iid}}{\sim} \text{Normal}(0,1)$ then $Y \sim \chi_k^2$.

Definition 0.8 - *Exponential Distribution*
Let $X \sim \text{Exponential}(\lambda)$.
A *continuous* random variable modelled by a *Exponential Distribution* with rate-parameter λ. Then

$$\begin{array}{rcl} f_X(x) &=& \mathbb{1}\{t\geq 0\}.\lambda e^{-\lambda x} \\ F_X(x) &=& \mathbb{1}\{t\geq 0\}.\left(1-e^{-\lambda x}\right) \\ \mathbb{E}(X) &=& \frac{1}{\lambda} \\ \text{Var}(X) &=& \frac{1}{\lambda^2} \\ \mathcal{M}_X(t) &=& \mathbb{1}\{t<\lambda\}\frac{\lambda}{\lambda-t} \end{array}$$

N.B. Exponential Distribution is used to model the wait time between decays of a radioactive source.

Definition 0.9 - *Gamma Distribution*
Let $X \sim \Gamma(\alpha,\beta)$.
A *continuous* random variable modelled by a *Gamma Distribution* with shape parameter $\alpha > 0$ & rate parameter β. Then

$$\begin{array}{rcl} f_X(x) &=& \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x} \\ F_X(x) &=& \frac{\Gamma(\alpha)}{\alpha}\gamma(\alpha,\beta x) \\ \mathbb{E}(X) &=& \frac{\alpha}{\beta} \\ \text{Var}(X) &=& \frac{\alpha}{\beta^2} \\ \mathcal{M}_X(t) &=& \mathbb{1}\{t<\beta\}\left(1-\frac{t}{\beta}\right)^{-\alpha} \end{array}$$

N.B. There is an equivalent definition of a *Gamma Distribution* in terms of a shape & scale parameter. The scale parameter is 1 over the rate parameter in this definition.

Definition 0.10 - *Multinomial Distribution*
Let $\mathbf{X} \sim \text{Multinomial}(n,\mathbf{p})$.
A *discrete* random varible which models n events with probability vector $\mathbf{p}$ for events $\{1,\dots,m\}$.

$$\begin{array}{rcl} f_{\mathbf{X}}(\mathbf{x}) & = & \mathbb{1}\left\{\sum_{i=1}^{m} x_i \equiv m\right\} \frac{n!}{x_1! \cdot \dots \cdot x_n!} \prod_{i=1}^{n} p_i^{x_i} \\ \mathbb{E}(X_i) & = & np_i \\ \text{Var}(X_i) & = & np_i(1-p_i) \\ \text{Cov}(X_i, x_j) & = & -np_ip_j \text{ for } i \neq j \\ \mathcal{M}_{X_i}(\theta_i) & = & \left(\sum_{i=1}^{m} p_i e^{\theta_i}\right)^n \end{array}$$

N.B. In a realisation $\mathbf{x}$ of $\mathbf{X}$, x_i is the number of times event i has occured.

Definition 0.11 - *Normal Distribution*
Let $X \sim \text{Normal}(\mu, \sigma^2)$.
A *continuous* random variable with mean μ & variance σ^2.

$$\begin{array}{rcl} f_X(x) & = & \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ F_X(x) & = & \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ \mathbb{E}(X) & = & \mu \\ \text{Var}(X) & = & \sigma^2 \\ \mathcal{M}_X(\theta) & = & e^{\mu\theta + \sigma^2\theta^2(1/2)} \end{array}$$

Definition 0.12 - *Pareto Distribution*
Let $X \sim \text{Pareto}(x_0, \theta)$.
A *continuous* random variable modelled by a *Pareto Distribution* with minimum value x_0 & shape parameter $\alpha > 0$. Then

$$\begin{array}{rcl} f_X(x) & = & \frac{\alpha x_0^\alpha}{x^{\alpha+1}} \\ F_X(x) & = & 1 - \left(\frac{x_0}{x}\right)^\alpha \\ \mathbb{E}(X) & = & \begin{cases} \infty & \alpha \leq 1 \\ \frac{\alpha x_0}{\alpha - 1} & \alpha > 1 \end{cases} \\ \text{Var}(X) & = & \begin{cases} \infty & \alpha \leq 2 \\ \frac{x_0^2 \alpha}{(\alpha-1)^2(\alpha-2)} & \alpha > 2 \end{cases} \\ \mathcal{M}_X(t) & = & \mathbb{1}\{t < 0\}\alpha(-x_0 t)^\alpha \Gamma(-\alpha, -x_0 t) \end{array}$$

Definition 0.13 - *Poisson Distribution*
Let $X \sim \text{Poisson}(\lambda)$.
A *discrete* random variable modelled by a *Poisson Distribution* with rate parameter λ. Then

$$\begin{array}{rcl} p_X(k) & = & \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for } k \in \mathbb{N}_0 \\ P_X(k) & = & e^{-\lambda} \sum_{i=1}^{k} \frac{\lambda^i}{i!} \\ \mathbb{E}(X) & = & \lambda \\ \text{Var}(X) & = & \lambda \\ \mathcal{M}_X(t) & = & e^{\lambda(e^t - 1)} \end{array}$$

N.B. Poisson Distribution is used to model the number of radioactive decays in a time period.

Definition 0.14 - *t-Distribution*
Let $X \sim t_r$.
A *continuous* random variable with r degrees of freedom. Then

$$\begin{array}{rcl}
f_X(k) & = & \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\dfrac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \\
\mathbb{E}(X) & = & \begin{cases} 0 & \text{if } \nu > 1 \\ \text{undefined} & \text{otherwise} \end{cases} \\
\text{Var}(X) & = & \begin{cases} \frac{\nu}{\nu-2} & \text{if } \nu > 2 \\ \infty & 1 < \nu \leq 2 \\ \text{undefined} & \text{otherwise} \end{cases} \\
\mathcal{M}_X(t) & = & \text{undefined}
\end{array}$$

N.B. Let $Y \sim \text{Normal}(0,1)$ & $Z \sim \chi^2_r$ be independent random variables then $X := \dfrac{Y}{\sqrt{Z/r}} \sim t_r$.

Definition 0.15 - *Uniform Distribution - Uniform*
Let $X \sim \text{Uniform}(a,b)$.
A *continuous* random variable with lower bound a & upper bound b. Then

$$\begin{array}{rcl}
f_X(x) & = & \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \\
F_X(x) & = & \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a,b] \\ 1 & \text{otherwise} \end{cases} \\
\mathbb{E}(X) & = & \frac{1}{2}(a+b) \\
\text{Var}(X) & = & \frac{1}{12}(b-a)^2 \\
\mathcal{M}_X(t) & = & \begin{cases} \dfrac{e^{tb}-e^{ta}}{t(b-a)} & t \neq 0 \\ 1 & t = 0 \end{cases}
\end{array}$$