Exam Notes

Andrew Martin

October 2018

Contents

1 Stochastic Decision Theory SDT

1.1 Formulae:

1.2 Proofs/Tricks:

1.2.1 HMM

Note most proofs here use some combination of:

- Law of Total Probability
- Definition of Conditional probability
- Markov Property
- Conditional Independence

1.3 Paraphrased Notes:

1.3.1 Stochastic Linear Programming

For general optimisation problems make sure to indicate variable meanings and give the maximum/minimum cost as well as the values which achieve these. Standard form:

$$\begin{aligned} \min\ & z = c^T x \\ \text{such that}\ & Ax = b \\ & x \geq 0 \end{aligned}$$

Where $b \geq 0$.

If we have inequalities

$$A'x \geq b$$

We introduce and subtract/add slack variables such that the slack variables $x_i \geq 0$.

The x which satisfy $Ax = b, x \geq 0$ are **feasible solutions**.

The *rank* is the number of linearly independent columns. The set of linearly independent columns will make a basis.

If B is a basis, then $Bx_B = b$ has a unique solution. Let $x := (x_B, \vec{0})$. Then x satisfies $Ax = b$ and is called a **basic solution** with respect to the basis B.

The feasible region is the **set of all feasibly solutions**. It can be

- Empty - no solutions
- Unbounded convex - infinite solutions
- Convex polytope - bounded number of solutions

Any point $z \in S$ where S is convex, which **cannot** be expressed as

$$\lambda x + (1 - \lambda)y$$

For $x, y \in S$ and $\lambda \in (0, 1)$, is an **extreme point**.
Extreme points include vertices and the entirety of edges on rounded shapes.

The **optimum** of a linear objective function with a convex polytope feasible region will occur at an extreme point.
A point in the feasible region is extreme iff it is a **basic feasible solution**.

The **simplex method** solves linear problems in std form by transforming from bases such that the objective function's value doesn't increase. `linprog()` uses this.

The **dual** problem of a standard linear program has form

$$\begin{aligned} \max\ & w = y^T b \\ \text{such that } & A^t y \leq c \\ & y \text{ free.} \end{aligned}$$

Weak duality states that the for feasible solutions of z, w for the primal and dual respectively

$$z \geq w$$

Strong duality states that **if** an optimal solution exists

$$\min z = \max w$$

If there is no optimal to the primal (or dual) then the primal and dual have no feasible solutions.
Table of conversions:

Primal	Dual
$\max z = \sum_{j=1}^{n} c_j x_j + z_0$	$\min w = \sum_{i=1}^{m} y_i b_i + z_0$
$\sum_{j=1}^{n} a_{ij} x_j = b_i$	y_i free
$\sum_{j=1}^{n} a_{ij} x_j \leq b_i$	$y_i \geq 0$
$\sum_{j=1}^{n} a_{ij} x_j \geq b_i$	$y_i \leq 0$
$x_j \geq 0$	$\sum_{i=1}^{m} y_i a_{ij} \geq c_j$
$x_j \leq 0$	$\sum_{i=1}^{m} y_i a_{ij} \leq c_j$
x_j free	$\sum_{i=1}^{m} y_i a_{ij} = c_j$

Table 1: Comparison of Primal and Dual constraints

When considering **stochastic LPs** we have one viable option - Recourse DEP (RDEP)- use expectations and 'recourse' variables
a RDEP will have form

To approximate the PMF of a random variable Z:

1. Generate samples $\mu = 1, \ldots, K$, of $z^{(\mu)}$ from Z restricted to the 99% CI
2. create equi-spaced partitions of the CI into r sub-intervals I_ν
3. Calculate the mean $\bar{z}_\nu$ of $z^{(\mu)} \in I_\nu$ for $\nu = 1, \ldots, r$ to estimate

$$E[Z_i | Z_i \in I_\nu]$$

4. Calculate for each I_ν the frequency p_ν which estimates

$$P(Z_i \in I_\nu)$$

The **empirical reliability** $\rho(\tilde{x})$ is the probability that $\tilde{x}$ is feasible s.t. the recourse variables are $= 0$.

$$P(A\tilde{x} \geq f(\tilde{x}, Z))$$

The empirical reliability is a measure of how likely it is that we won't have recourse - it will not match the theoretical reliability due to disretisation errors.

Recourse: For $i = 1, \ldots, m$ and for each realisation ϵ of the random variable

$$g_i^+(x, \epsilon) = \begin{cases} 0 & g_i(x, \epsilon) \leq 0 \\ g_i(x, \epsilon) & otherwise \end{cases}$$

This is a measure of how much recourse we have to pay for. I.e. if $g_i^+ > 0$ for some x, ϵ, then we have violated the i^{th} constraint, and must pay recourse.

In an RDEP, if this occurs, we have to take recourse $y_i(\epsilon)$ to satisfy

$$y_i(\epsilon) \geq g_i^+(x, \epsilon)$$

I.e. to pay recourse such that there is no constraint violation (pizza example).

General **linear** recourse program. Let $\vec{y} \in \mathcal{Y}$ be the recourse variables, where $\mathcal{Y}$ is the constraint polytope and with unit cost vector $\vec{q}$ and W the linear function for the recourse. This gives the recourse function

$$Q(x, \epsilon) = \min_y \{q^T y(\epsilon) : Wy(\epsilon) \geq g^+(x, \epsilon), y(\epsilon) \in \mathcal{Y}\}$$

General **non-linear** recourse program where $H_i(y)$ is a given recourse function

$$Q(x, \epsilon) = \min_y \{q(y) : H_i(y) \geq g_i^+(x, \epsilon)\}$$

Two-stage stochastic recourse program. If it is meaningful to minimise expectations we use:

$$\min_x E_\zeta[f_0(x, \zeta)] = \min_x E_\zeta[g_0(x, \zeta) + Q(x, \zeta)]$$

Stochastic linear program w recourse

$$\begin{aligned} \min\ & w = c^T x \\ \text{such that}\ & Ax = b \\ & T(\zeta)x = h(\zeta) \\ & x \geq 0 \end{aligned}$$

And we assume

$$T(\zeta) = \hat{T}_0 + \sum_{i=1}^{k} 1_{\epsilon_i} \hat{T}_i$$

And h is the same but replace T with h. And we assume that $\hat{T}$ is deterministic, and 1_{ϵ_i} is the indicator function.

Two stage-stochastic linear program w **fixed** recourse

$$\begin{aligned} \min\ & w = c^T x + E_\zeta[Q(x, \zeta)] \\ & \text{such that}\ Ax = b \\ & x \geq 0 \end{aligned}$$

Where for each realisation ϵ of ζ:

$$Q(x, \epsilon) = \min_y \{q^T y : Wy = h(\epsilon) - T(\epsilon)x, y \geq 0\}$$

Expanded DEP form: For the two-stage program where ζ is finite with strictly positive probabilities we can write:

$$\begin{aligned} \min\ & w = c^T x + \sum_{k=1}^{r} p_k q^T y_k \\ & \text{such that}\ Ax = b \\ & T(\epsilon_k)x + Wy_k = h(\epsilon_k) \\ & x, y_k \geq 0 \end{aligned}$$

(We will talk about this program for awhile)

The fixed recourse is **complete** if the $m_1 \times \bar{n}$ matrix W satisfies:

$$\{z : z = Wy, y \geq 0\} = \mathbb{R}^{m_1}$$

Where $\bar{n}$ is the number of recourse variables and m_1 is the number of stochastic constraints.
I.e. the second stage program will always be feasible.
If $g_0(x,\epsilon)$ and $Q(x,\epsilon)$ are convex for x in each ϵ, and X is a convex set, then the two-stage stochastic recourse program is a **convex** program.
Convexity: For a function $f(x)$ to be convex it must satisfy Jensen's inequality

$$E[h(x)] \leq h(E[x])$$

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

The Second last line generates the induced feasibility set. I.e. the Induced Feasibility set is the set K such that:

$$K = \{x : T(\epsilon_j)x + Wy_j = h(\epsilon_j), y_j \geq 0, j = 1, \ldots, r\}$$

If the set $\epsilon \in S_\zeta$ is either finite or a convex polytope, then K is a convex polytope. The feasible set of x is then restricted to

$$x \in X \cap K$$

The $m_1 \times \bar{n}$ matrix, W is a complete recourse matrix if:
$Rank(W) = m_1$ And assuming the first m_1 cols of W are indep, then:

$$\begin{aligned} Wy &= 0 \\ y_i \geq 0, i &= 1, \ldots, m_1 \\ y &\geq 0 \end{aligned}$$

Must have a feasible solution

If $K \cap X = X$ then the program has relatively complete recourse
A Chance constrained program is defined as

$$\begin{aligned} \min \quad & w = E_\zeta[f_0(x,\zeta)] \\ \text{such that } & P(\epsilon : g_i(x,\epsilon) \leq 0) \geq a_i \end{aligned}$$

I.e. such that the probability of recourse being required is below a_i.
If $g(x,\epsilon)$ is jointly convex, and the probability measure P is quasi-concave, then for all $\alpha \in [0,1]$,

$$B(\alpha) = \{x : P(\{\epsilon : g(x,\epsilon) \leq 0\}) \geq \alpha\}$$

Is convex.
Quasi-concave means:

$$P(\lambda S_1 + (1-\lambda)S_2) \geq \min\{P(S_1), P(S_2)\}$$

For all $\lambda \in (0,1)$, and convex sets S_i.

Lagrange multipliers: Define the lagrangian as:

$$\Delta L(\vec{x}, \vec{p}) = f(\vec{x}) + p^T g(Ax + b)$$

Where $\vec{p}$ is a vector of Lagrange Multipliers. Solving the system:

$$\min f(x) \quad \text{such that } Ax + b \leq 0$$

Is equivalent to

$$\min_x \left(\max_{\vec{p}} L(\vec{x}, \vec{p}) \right)$$

The two-stage recourse DEP has dual-decomposition structure. It can be solved using the Dual-Decomposition method:

1. Denote
$$B_0 := \{(x^T, \theta)^T : Ax = b, x \geq 0, \theta \in \mathbb{R}\}$$
And Let θ_0 be a lower bound for
$$\min_y \{q^T y : Ax = b, Yx + Wy = h, x, y \geq 0\}$$
And solve
$$\min_{x,\theta} \{c^T x + \theta : Ax = b, \theta \geq \theta_0, x \geq 0\}$$
And denote the solution $(\hat{x}, \hat{\theta})$. Let the set $B_1 := \{R^n \times \{\theta\} : \theta \geq \theta_0\}$

2. solve:
$$\begin{aligned} Q(\hat{x}) &= \min\{q^T y : Wy = h - T\hat{x}, y \geq 0\} \\ &= \max\{(h - T\hat{x})^T u : W^T y \leq q, u \text{ free}\} \end{aligned}$$
If $Q(\hat{x}) \leq \hat{\theta})$ stop.
If $Q(\hat{x})$ is bounded, we have to take an optimality cut, and redefine
$$B_1 := B_1 \cap \{(x, \theta) : \theta \geq \hat{u}^T(h - Tx)\}$$
And continue to step 3. If $Q(\hat{x})$ unbounded, cut the infeasible solution $\hat{x}$: Such that the set becomes
$$B_1 := B_1 \cap \{(x, \theta) : \tilde{u}^T(h - Tx) \leq 0\}$$
Proceed to step 3

3. Solve the updated LP:
$$(\tilde{x}, \tilde{\theta}) = \min\{c^T x + \theta : (x^T, \theta) \in B_0 \cap B_1\}$$
And return to step 2 with $(\hat{x}, \hat{\theta}) := (\tilde{x}, \tilde{\theta})$

If the program is solvable and the set B_0 is bounded, then this method converges after finitely many steps. The L-shaped algorithm solves these for harder stochastic problems:
Initially set
$$t = s = v = 0,\ \vec{D}_0 = 0,\ d_0 = 0,\ \vec{E}_0 = 0,\ e_0 = 0,\ \theta = 0$$

1. Increment v (iteration counter) and solve
$$\begin{aligned} \min\ z = c^T x + \theta \\ \text{such that } Ax = b \\ D_l x \geq d_l, \quad l = 1, \ldots, t \\ E_l x + \theta \geq e_l, \quad l = 1, \ldots, s \\ x \geq 0, \theta \in R \end{aligned}$$
Let $(x^{(v)}, \theta^{(v)})$ denote the solution, and set $\theta^{(v)} = -\infty$ when $s = 0$

2. For $k = 1, \ldots, r$, while $\tilde{w} > 0$ solve:
$$\begin{aligned} \min\ \tilde{w} = e^T v^+ e^T v^- \\ \text{such that } Wy_k^{(v)} + v^+ - v^- = h_k - T_k x^{(v)} \\ y_k^{(v)} \geq 0, v^+, v^- \geq 0 \end{aligned}$$
Where e^T is a vector of ones. (i.e. we sum v^+ and v^-).If for any k, we get $\tilde{w} > 0$, let $\sigma_k^{(v)}$ be the vector of Lagrange multipliers (obtained by solving the dual), and define:
$$\begin{aligned} \vec{D}_{l+1} &= (\sigma^{(v)})^T T_k \\ d_{l+1} &= (\sigma^{(v)})^T h_k \end{aligned}$$
Increment t and go to step 1 Else if $\tilde{w} = 0$ for all k, Then a feasible solution exists, and we can proceed to step 3.

3. For $k = 1, \ldots, s$ solve:

$$\begin{aligned} \min \hat{w} &= q^T y_k^{(v)} \\ \text{such that } W y_l^{(v)} &= h_k - T_k x^{(v)} \\ y_k^{(v)} &\geq 0 \end{aligned}$$

Let $\pi_k^{(v)}$ be the vector of Lagrange multipliers associated with an optimal solution for a given k and define

$$\begin{aligned} E_{s+1} &= \sum_{k=1}^{r} p_k (\pi_k^{(v)})^T T_k \\ e_{s+1} &= \sum_{k=1}^{r} p_k (\pi_k^{(v)})^T h_k \end{aligned}$$

And set

$$f^{(v)} = e_{s+k} - E_{s+1} x^{(v)}$$

If $\theta^{(v)} \geq f^{(v)}$ then stop the iteration procedure as $x^{(v)}$ is optimal
Otherwise increment s and return to 1

1.3.2 Markov Models

For a valid markov model we must have points where we can make decisions, distinct finite states, actions, rewards/costs, and transition probabilities. We assume the markov properties exist.

Finite horizon Problems: Want to find the optimal set of policies.
The policy Ψ has a markov process X_n^Ψ on the state space S for $n = 0, \ldots, T$ with transition probabilities:

$$P(X_{n+1}^\Psi = j | X_n^\Psi = i) = p_{ij}(f_n(i))$$

The expected cost of the policy Ψ, denoted J^Ψ is

$$J_k^\Psi(i) = \mathbb{E}\left[C(X_T^\Psi) + \sum_{n=k}^{T-1} c(X_n^\Psi, f_n(X_n^\Psi)) \right]$$

Where $X_k^\Psi = i$, and c is the cost function.
This is essentially saying

$$\text{Expected cost} = \text{final cost} + \text{Cost accumulated until the last step}$$

The optimal policy, $\Psi^* = \{f_0^*, f_1^*, \ldots, f_{T-1}^*\}$ satisfies:

$$J^{\Psi^*}(i_0) = \min_{\Psi \in \Gamma} J_0^\Psi(i_0)$$

Where Γ is the set of all admissible policies.

The truncated policy (by removing the first k steps) is optimal for the sub-problem. I.e. If $\{f_0^*, \ldots, f_{T-1}^*\}$ is optimal for the problem, then $\{f_k^*, \ldots, f_{T-1}^*\}$ is optimal for the sub-problem.
Bellman Optimisation principle uses this - start with $J_T(i) = C(i)$ (where $C(i)$ is the terminal cost for state i) and iterate backwards through k

$$J_k(i) = \min_{u_k \in U_k(i)} \mathbb{E}\left[c(i, u_k) + \sum_{j \in S} p_{ij}(u_k) J_{k+1}(j) \right]$$

Then the set of u_k which minimises this generates the optimal policy.

Indefinite Horizon problems: restrict policies to stationary policies.

Shortest path MDP with finite state space $i = 1, \ldots, N$

$$\mathbb{E}[c(i,u)] = \sum_{j=1}^{N} p_{ij}(u) r_{ij}$$

Where r_{ij} is the cost/reward for the transition $i \to j$

Average Value MDP Consider the stationary distribution π such that $\pi P = \pi$. So we get the expected reward for state i is

$$q_i = \sum_{j=1}^{n} p_{ij} r_{ij}$$

And the gain is (total expected reward)

$$g = \sum_{i=1}^{n} \pi_i q_i$$

I.e. the expected gain is the expected reward for that state times the time spent being in that state

By using policy iteration we come to the Policy Improvement Routine (PIR):
Pick a policy, and solve:

$$\phi(i) + g = \left(\sum_{j=1}^{n} p_{ij} r_{ij} \right) + \left(\sum_{j=1}^{n} p_{ij} \phi(j) \right)$$

for all ϕ and g where we set $\phi(1) = 0$
Then find an alternative policy f' and see if it improves

$$\left(\sum_{j=1}^{n} p'_{ij} r'_{ij} \right) + \left(\sum_{j=1}^{n} p'_{ij} \phi(j) \right)$$

NOTING THAT WE USE THE SAME $\phi(j)$
If it improves that function, we set $p_{ij} = p'_{ij}$ and $r_{ij} = r'_{ij}$ and recalculate ϕ and g.

Infinite Horizon problems have a positive recurrent Markov Chain with (effectively) infinite termination time. The average cost per stage g is defined as:

$$g = \lim_{T \to \infty} \frac{1}{T} \mathbb{E} \left[\sum_{n=0}^{T} v(X_n, f_n(X_n)) \right]$$

And $-\infty < g < \infty$

Discounted MDPs occur when we consider a timed discount factor $\beta \in (0, 1]$ such that, for example the cost matters more to us now than later. We define the cost as

$$\sum_{n=0}^{T-1} \beta^n c(X_n, u_n) + \beta^T C(X_t)$$

We then find that if the costs are all bounded by a constant B, i.e. $|c(X_n, u_n)| < B$ for all X_n, and $\beta \in (0, 1)$, the infinite horizon cost is bounded by $B/(1-\beta)$ (geometric sum thing) The Minimal present value $J_k(i)$ is:

$$\inf_{\psi} \mathbb{E} \left[\sum_{n=k}^{T-1} c(X_n^{\psi}, f_n(X_n^{\psi})) + \beta^{T-k} C(X_T) \;\middle|\; X_k = i \right]$$

Giving the optimality equations

$$J_k(i) = \inf_{u_k \in U_k(i)} \{ \mathbb{E}[c(i, u_k) + \beta J_{k+1}(X_{k+1})] \}$$

With $J_T(i) = C(X_T)$

Negative/Positive/Discounted Programming (N/P/D)
Discounted (D) if $\beta \in (0, 1)$ and $|c(i,u)| < B$

Negative (N) if $\beta \in (0,1]$ and $c(i,u) \geq 0$ (i.e. we expect to lose money)
Positive (P) if $\beta \in (0,1]$ and $c(i,u) \leq 0$ (i.e. we expect to gain money)

For N and P we usually take $\beta = 1$ for simplicity.

Let assumption A be the assumption that there are finitely many actions that can be taken (U is a finite set) in each state.

If D or P holds, or N and A hold together, then the optimal infinite horizon cost $F_\infty(i) = F(i)$

Negative Programming and Optimal Stopping - problem D or problem N with A guarantees an optimal policy.

Optimal Stopping over finite horizon - if there is an absorbing state such that $u = 0$ to stop and $u = 1$ to continue, and in this absorbing state there is no further cost. I.e. we have

$$c(x,u) = \begin{cases} k(x), & u = 0 \\ c(x), & u = 1 \end{cases}$$

Let $F_s(x)$ denote the minimum total cost when we have to stop in the next s steps. We get

$$F_s(x) = \min\{k(x), c(x) + \mathbb{E}[F_{s-1}(X_1)|X_0 = x, u_0 = 1]\}$$

Instead can write this as a set

$$S = \{x : k(x) \leq c(x) + \mathbb{E}[k(X_1)|X_0 = x, u_0 = 1]\}$$

And we want to stop when $x \in S$.

Optimal Stopping over infinite horizon - if costs are bounded, then S works for this as well!

Hidden Markov Models - a model where there is an unobserved underlying process $\{X_t\}_{t\geq 0}$ with an observed process $\{Y_t\}_{t\geq 0}$.
The hidden markov model $\{(X_n, Y_n), n \geq 0\}$ gives:

$$P(\vec{y}_t|\vec{x}_t) = \prod_{i=0}^{t} P(y_i|x_i)$$

I.e. $Y_i|X_i$ is conditionally independent of $X_0, \ldots, X_{i-1}$. Let $\Lambda = (\theta_p, \phi_p, \vec{p}_0)$ where:
θ_p is the p_{ij} for X,
ϕ_p is $P(y|x)$, and,
p_0 is $P(X_0)$

Problem 1: Given parameters and observed sequence, find probability of that observed sequence, i.e. find

$$P(\vec{y}|\Lambda) = \sum_x P(x, y|\Lambda)$$

Forward Function/Algorithm

1. $i = 1, \ldots, N$:
$$\alpha_0(i) = P(y_0, X_0 = i, \Lambda) = P(X_0 = i)P(y_0|X_0 = i, \Lambda)$$
2. For $t = 1, \ldots, T$ and $i = 1, \ldots, N$
$$\alpha_t(i) = \left(\sum_{j=1}^{N} \alpha_{t-1}(j)p_{ji}\right) P(y_t|X_t = i)$$
3. The final answer is
$$P(\vec{y}_t|\Lambda) = \sum_{i=1}^{N} \alpha_t(i)$$

Problem 2: Given parameters and observed sequence, find most likely hidden sequence I.e.

$$P(X|y,\Lambda)$$

Backward Function/Algorithm
Define

$$\beta_t(i) = P((y_{t+1},\dots,y_T)|X_t = i,\Lambda)$$

1. $\beta_T(i) = 1$, for $i = 1,\dots,N$
2. $t = T-1, T-2,\dots,0$ and $i = 1,\dots,N$

$$\beta_t(i) = \sum_{j=1}^{N} P(y_{t+1}|X_{j+1} = j,\Lambda)\beta_{t+1}(j)$$

Forward/backward algorithm:

$$\gamma_t(i) := P(X_t = i|y,\Lambda) = \frac{\alpha_t(i)\beta_t(i)}{P(y|\Lambda)}$$

Viterbi algorithm

1. For $i = 1,\dots,N$ set

$$\delta_0(i) = p_0(i)P(y_0|X_0 = i)$$

2. For $t = 1,\dots,T$ and $i = 1,\dots,N$

$$\delta_t(i) = \max_{j=1,\dots,N}(\delta_{t-1}(j)p_{ji}P(y_t|X_t = i))$$

Such that the most likely path at time t is i which maximises $\delta_t(i)$

Log-Viterbi algorithm - to make up for arithmetic errors for small calculations: It is literally the log of the viterbi algorithm.

1. For $i = 1,2,\dots,N$ let

$$\hat{\delta_0}(i) = \log(p_0(i)) + \log(P(y_0|X_0 = i))$$

2. For $t = 1,2,\dots,T$ and $i = 1,2,\dots,N$, compute $\hat{\delta}_t(i)$ given by

$$\max_{j=1,2,\dots,N}\left[\hat{\delta}_{t-1}(j) + \log(p_{ji}) + \log(P(y_t|X_t = i))\right]$$

Problem 3: Given observed sequence and dimension of Λ, find Λ^* which maximises probability of obtaining the observed sequence
Baum-Welch Algorithm

1. Initialise $\Lambda = (\theta_p, \phi_p, \vec{p_0})$ by guessing (IT CANNOT BE EXACTLY UNIFORM)
2. Using Λ compute

$$\alpha_t(i), \beta_t(i), \gamma_t(i), \zeta_t(i,j)$$

3. Reestimate Λ':
   (a) For $i = 1,\dots,N$

$$p_0(i) = \gamma_0(i)$$

   (b) For $i = 1,\dots,N$ and $j = 1,\dots,N$

$$p_{ij} = \sum_{t=0}^{T-1}\zeta_t(i,j) \Big/ \sum_{t=0}^{T-1}\gamma_t(i)$$

(c) For $j = 1, \ldots, N$ and $k = 1, \ldots, M$

$$P(k|j) = \sum_{y_t=k, t=0}^{T-1} \gamma_t(j) \Big/ \sum_{t=0}^{T-1} \gamma_t(j)$$

4. If

$$P(\vec{y}|\Lambda') > P(\vec{y}|\Lambda)$$

Set $\Lambda = \Lambda'$
Otherwise, stop.

(EM) Expectation-Maximisation Algorithm

- y observed data
- x unobserved data (hidden)
- θ parameters

Find the MLE $\theta^* = \arg\max_\theta \log f(y|\theta)$

1. Set $n = 0$ and choose a starting θ_0
2. Calculate

$$E[\log f(\vec{y}, \vec{x}|\theta)|\vec{y}] =: Q(\theta, \theta_n)$$

3. Maximise: Set $n = n + 1$ and let

$$\theta_n := \arg\max_{\theta \in \Theta}(\theta, \theta_{n-1})$$

It is unlikely we will be asked to use the EM algorithm.

2 Random Processes RP

Rough Algorithm for simulating CTMC:

1. Given current state S
2. Compute rate of each event
3. Sum them to get he total rate (TR) of an event
4. Sample an exponentially distributed time with mean $\frac{1}{TR}$
5. Choose the event to occur with probability $\propto$ its rate, e.g. death of a susceptible would be chosen with probability $\frac{\mu * S}{TR}$
6. Update the state given the chosen event

Markov property:

$$P\left(X(t+s)=k|X(u)=i, X(s)=j, u<s)\right) = P(X(t+s)=k|X(s)=j)$$

For all $s,t \geq 0$ and $i,j,k \in S$

Let $P(a,b)$ be the matrix of transition probabilities in the interval (a,b).
The CTMC is time homogeneous if

$$P(s,t+s) = P(0,t) := P(t)\ \forall t, s \geq 0$$

For the transition probability matrix, the elements (the transition probabilities) must obey:

$$P_{ij}(t) \geq 0$$
$$\sum_{j \in S} P_{ij}(t) = 1$$

We will assume these hold. The second is the 'honesty' condition.

Let the sequence $t_1, t_2, t_3, \ldots$ be iid $Exp(\lambda)$ random variables, with rate λ

$$T_0 := 0$$
$$T_n := \sum_{i=1}^{n} t_i$$
$$N(t) := max(n : T_n \leq t), \quad t \geq 0$$

Then $(N(t), t \geq 0)$ is called the poisson process.

Think of T_n as the arrival time of the n^{th} student, and $N(t)$ as the number of arrivals by time t. Note the t_n refer to the **time steps**.

The sum $T_n = t_1 + t_2 + \ldots + t_n$ where $t_i \sim Exp(\lambda)$ has an Erlang distribution $Erlang(n, \lambda)$, with pdf:

$$f_{T_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n+1)!} \quad for \quad t \geq 0$$

Consider a CTMC X on the state space S, with generator Q. Then for all $i \in S$

$$P(\text{moves to state } j \neq i | \text{ leaves state } i \text{ at time } t) = \frac{q_{ij}}{-q_{ii}}$$

The minimum of a group of iid $T_i \sim Exp(\lambda_i)$, $i = 1, \ldots, n$ Let $M := \min\{T_1, T_2, \ldots, T_n\}$

$$\implies M \sim Exp(\sum_{i=1}^{n} \lambda_i)$$

The time that the system stays in state i is exponentially distributed, with rate $-q_{ii}$.

Recall first principles of the derivative:

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=f'(x)$$

The generator Q of a CTMC X has entries (when the limit exists):

$$q_{ij}=\lim_{h\to 0^+}\frac{P_{ij}(h)-P_{ij}(0)}{h}=\frac{P_{ij}(h)}{h}\quad(\geq 0), i\neq j, j\in S$$
$$q_{ii}=\lim_{h\to 0^+}\frac{P_{ii}(h)-P_{ii}(0)}{h}=\frac{P_{ii}(h)-1}{h}\quad(\leq 0)$$

Where $P_{ij}(h):=P(X(t+h)=j|X(t)=i)$ is the conditional probability that the system is in j by the end of time interval h.
Loosely, Q is the right-derivative of the matrix $P(t)$ at the point $t=0$. In matrix notation we can rewrite this as

$$Q=\lim_{h\to 0^+}\frac{P(h)-I}{h}$$

Where the rows sum to 0 (conservative):

$$\sum_{j\in S}q_{ij}=0$$

Recall Eulers method:

$$f(x+h)\approx f(x)=hf'(x)$$

q_{ij} is the *instantaneous rate* that the chain moves from $i\to j$, and $-q_{ii}$ is the instantaneous rate that the chain moves out of state i

Chapman-Kolmogorov Equation For a CTMC $X=(X(t),t\geq 0)$ on a state space S, with $i,j\in S$, we have

$$P_{ij}=\sum_{k\in S}P_{ik}(u)P_{kj}(t-u),\quad 0<u\leq t$$

Or in matrix notation

$$P(t)=P(u)P(t-u)$$

The Kolmogorov Backward Differential Equations for a CTMC,

$$\frac{\partial P_{ij}(t)}{\partial t}=\sum_{k\in S}q_{ik}P_{kj}(t)$$

Or in matrix notation

$$\frac{\partial P(t)}{\partial t}=QP(t)$$

Kolmogorov Forward Differential Equations for a CTMC

$$\frac{\partial P_{ij}(t)}{\partial t}=\sum_{k\in S}P_{ik}(t)q_{kj}$$

The solution to the Kolmogorov differential equations is

$$P(t)=e^{Qt},\quad\text{where } e^A=\sum_{n=0}^{\infty}\frac{A^n}{n!}$$

Spectral Representation of Q. Assume the eigenvalues of Q are all distinct (not necessary but makes it easier) and labelled λ_i, we have

$$Q=RDL=\lambda_1M_1+\lambda_2M_2+\ldots+\lambda_dM_d$$

Where $D = diag(\lambda_1, \lambda_2, \ldots, \lambda_d)$, R has columns $r_1', r_2', \ldots, r_d'$ L has rows $l_1, l_2, \ldots, l_d$ and so $M_i = r_i' l_i$ with $l_i(1 \times d)$ the left eigenvector of Q corresponding to the eigenvalue λ_i and $r_i'(d \times 1)$ the right eigenvector of Q corresponding to the eigenvalue λ_i such that $r_i l_j' = \delta_{ij}$ Hence we get

$$P(t) = \exp(Qt) = e^{\lambda_1 t} M_1 + e^{\lambda_2 t} M_2 + \ldots + e^{\lambda_d t} M_d$$

The generating function $P(z,t)$ for a process with transition probabilities $p_{0n}(t)$ for $n = 0, 1 \ldots$ is given by

$$P(z,t) = \sum_{n=0}^{\infty} P_{0n}(t) z^n$$

Let $N(t)$ be the r.v. representing the number of events by time t. The expectation will be

$$E[N(t)] = \sum_{j=0}^{\infty} j P(N(t) = j)$$

We can also calculate these moments using the KBDEs/KFDEs
Sometimes we can't use the KBDE/BFDEs (because each time we do it it depends on a higher order moment).

Stationary distribution of a CTMC $X(t)$:

$$\pi = \pi P(t) \forall t \geq 0$$

Or

$$\pi Q^n = 0 \forall n \geq 1$$

For this course the equilibrium distribution is the solution to the stationary distribution.
Global Balance Equations (flux in = flux out)

$$\pi_j \sum_{\substack{k \in S \\ k \neq j}} q_{jk} = \sum_{\substack{k \in S \\ k \neq j}} \pi_k q_{kj}$$

Hitting Probabilities derive using a first step analysis. The probability f_i that this CTMC ever reaches state j, given it starts in state i is given by the *minimal non-negative solution* to the equations

$$(-q_{ii}) x_i = \sum_{\substack{k \in S \\ k \neq i}} q_{ik} x_k$$

Subject to the boundary condition $x_j = 1$.
Expected Hitting Times

$$t_i = -\frac{1}{q_{ii}} + \sum_{\substack{k \neq i \\ k \in S}} \frac{q_{ik}}{-q_{ii}} t_k, \quad \text{with } t_j = 0$$

Note you can think of $-\frac{1}{q_{ii}}$ as the expected time until next transition
$\frac{q_{ik}}{-q_{ii}}$ as the probability of jumping to state k at the next transition
and t_k as the expected time to reach j given the process starts in state k
A CTMC $\{X(t)\}$ is stationary if

$$P(X(t_1) = i_1, \ldots, X(t_n) = i_n) = P(X(t_1 + \delta) = i_1, \ldots, X(t_n + \delta) = i_n)$$

For every positive integer n and for all t_j, δ.
reversed-time CTMC $\{X^R(t)\}$ to be

$$X^R(t) = X(\tau - t), \text{ for arbitrary } \tau$$

If $X(t)$ is a stationary CTMC with transition rates q_{jk} and equilibrium distribution π, then the reversed-time process $X^R(t) = X(\tau - t)$ is a stationary CTMC with transition rates

$$q_{kj}^R = \frac{q_{jk} \pi_j}{\pi_k}$$

for $k, j \in S$ and $k \neq j$. The equilibrium distribution will be $\pi_j^R = \pi_j$, and

$$q_{jj} = q_{jj}^R$$

For a M/M/1 queue The equilibrium probabilities are

$$\pi_j = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^j$$

For $\lambda < \mu$.

A CTMC is reversible if the reversed-time process has the same transition rates as the forward time process. I.e.

$$q_{jk}^R = q_{jk}, \quad \forall j, k \in S$$

Note: Clear distinction between a reversed-time process and a reversible process. All **stationary** CTMCs may be looked at in reverse time, but **all** CTMCs are reversible.

A stationary CTMC is **reversible** if and only if the **detailed balance equations** are satisfied:

$$\pi_j q_{jk} = \pi_k q_{kj}$$

These will be the equilibrium distribution of the markov chain.
Assume reversibility then find $\pi_j > 0$ that satisfy the equations. If it works, we have found the equilibrium distribution of the CTMC, and it is reversible.

A stationary k-variate birth-and-death process has equilibrium distribution

$$\pi(n) = C \prod_{r=1}^{k} \prod_{k=1}^{n_r} \frac{\lambda_r(l-1)}{\mu_r(l)}, \qquad \textit{where } C \textit{ is a normalising constant}$$

Consider 2 indep single server queues with arrival rates λ_i and service rates μ_i for $i = 1, 2$. These queues share a common waiting room of size R and customers which arrive to a full waiting room are lost. The state space of this system can be represented as follows:

1. Letting $[n]^+ = \max(0, n)$, we can write the state space A as:

$$A = \{(n_1, n_2) : [n_1 - 1]^+ + [n_2 - 1]^+ \leq R\}$$

2. Consider R to be ∞ which implies the two queues are totally independent, reversible birth/death processes and have joint equilibrium probability distribution given by (its gross):

$$\pi(n_1, n_2) = \left(1 - \frac{\lambda_1}{\mu_1}\right)\left(1 - \frac{\lambda_2}{\mu_2}\right)\left(\frac{\lambda_1}{\mu_1}\right)^{n_1}\left(\frac{\lambda_2}{\mu_2}\right)^{n_2}$$

   Assuming $\lambda_1 < \mu_1$ and $\lambda_2 < \mu_2$.

3. Since it was reversible, the truncated process must also be reversible with the same invariant measure. With the state space A as before, the equilibrium distribution $\bar{\pi}$ is given by

$$\bar{\pi}(n_1, n_2) = D\left(\frac{\lambda_1}{\mu_1}\right)^{n_1}\left(\frac{\lambda_2}{\mu_2}\right)^{n_2}$$

   Where

$$D = \left(\sum_A \left(\frac{\lambda_1}{\mu_1}\right)^{n_1}\left(\frac{\lambda_2}{\mu_2}\right)^{n_2}\right)^{-1}$$

Arrivals of packets (calls) λ_i which have their own destinations, D_i. Each destination D_i has capacity N_i. They first arrive at a common origin, where they are sent to their destination. The can alternatively get sent to a switch (which as M capacity), and then be sent to the destinations via a route with T_i capacity. Wow thats a shitty description.

Mean call holding times: exp with $1/\mu_i$ for each i.

First try the infinite case where the $N_i = \infty$ for all i. So effectively this means ignoring the whole switch part. This gives us r indep b/d processes. So the invariant measure is (letting n_i be the number of customers in the system from stream i.)

$$\prod_{i=1}^{r} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i} \frac{1}{n_i!}$$

The number of overflow circuits from the i^{th} direct route is then

$$[n_i - N_i]^+$$

Recalling that $[x]^+ = \max(x, 0)$
Therefore the state space A is restricted by:

$$[n_i - N_i]^+ \leq M \quad \& \quad [n_i - N_i]^+ \leq T_i$$

We can truncate the state space to A, giving

$$\pi(n_1, n_2, \ldots, n_r) = C \prod_{i=1}^{r} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i} \frac{1}{n_i!}$$

Where

$$C = \left(\sum_{A} \prod_{i=1}^{r} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i} \frac{1}{n_i!}\right)^{-1}$$

Arrivals at multiple origins, passing through a network with a bunch of redirects Assumptions:

- Fixed routing is used (no overflow/alternative)
- Finite number of links J inside the circuit, link j has c_j circuits, for $1 \leq j \leq J$
- R routes in total - calls requesting (combinations of origins & destinations), at route r arrive in a Poisson stream with rate λr, for $1 \leq r \leq R$
- WLOG, call holding times have unit mean.
- Route r calls uses $A_{j,r}$ circuits upon link j.

The matrix A has column as routes, and rows corresponding to which links are used.

Let n_r be the number of calls using route r and let $n = (n_1, \ldots, n_R)$ be the state of the network, let $c = (c_1, c_2, \ldots, c_J)$ be the circuits for each link. So $A = \{A_{j,r}\}$ be the $J \times R$ matrix. Then we can write

$$\begin{aligned} S(c) &= \{n : \sum_{r} A_{j,r} n_r \leq c_j, 1 \leq j \leq J\} \\ &= \{n : An \leq c\} \end{aligned}$$

Let c be infinite, giving an R variate b/d process, which is reversible!
Then, by truncating the process by making each c_j finite, gives

$$\pi(n) = [G(c, R)]^{-1} \prod_{r=1}^{R} \left(\frac{\lambda_r^{n_r}}{n_r!}\right), \quad for\ n \in S(c)$$

Where $c = (c_1, \ldots, c_J)^T$ and $S(c) = \{n : A_n \leq c\}$. $[G(c, R)]^{-1}$ is the normalising constant, which is dependent on the state space $S(c)$.

$$G(c, R) = \sum_{\{n : An \leq c\}} \prod_{r=1}^{R} \frac{\lambda_r^{n_r}}{n_r!}$$

Offered load is effectively λ_i/μ_i Erlangs

If $B_r := P(\text{call on route } r \text{ is blocked})$, and $e_r :=$ the r elementary vector or whatever it was called. Then

$$B_r = 1 - \frac{G(c - Ae_r, R)}{G(c, R)}$$

Erlang Loss System - the M/M/N/N queue
Arrival Process / Service Time Distribution / Number of servers / Capacity

The arrival rates

$$\lambda_l = \begin{cases} \lambda & 0 \leq l < N \\ 0 & l = N \end{cases}$$

Service rate:

$$\mu_l = l\mu \quad l \leq N$$

Equilibrium

$$\pi_i = \left[\sum_{l=0}^{N} \frac{1}{l!} a^l\right]^{-1} \frac{1}{i!} a^i$$

Where $a = \frac{\lambda}{\mu}$
The Erlang B formula :

$$B(N, a) = \pi_N = \frac{1}{N!} a^N \left[\sum_{l=0}^{N} \frac{1}{l!} a^l\right]^{-1}$$

Iterative version:

$$B(N, a) = \frac{aB(N-1, a)}{N + aB(N-1, a)}, \quad B(0, a) = 1$$

Erlang Fixed Point Method Assuming the links are **independent**, and calls are only accepted if accepted on all links In order to use the EFPM to determine blocking probabilities (in any nature), the following numerical procedure can be invoked:

1. Initially assume $\alpha = (\alpha_j, j \in J) = \vec{0}$
2. Calculate $y = (y_j, j \in J)$ using the current value of α

$$y_j = \sum_{r: j \in r} a_r \prod_{\substack{i \in r \\ i \neq j}} (1 - \alpha_i)$$

3. Use that to calculate α using the Erlang B formula

$$\alpha_j = B(c_j, y_j) \quad \text{(Erlang loss formula)}$$

4. Compare the estimated $\alpha's$ to the previous $\alpha's$ based on some tolerance.
5. If they don't conform to that tolerance, repeat steps 2-4
6. Calculate $B = (B_r, r \in R)$ according to its formula (this is not the erlang B!)
7. Terminate. α is close enough as far as we care.

Burke's Theorem: Consider a queue with a Poisson arrival process with rate λ, and exponential service time distribution with parameter $\mu > \lambda$. In equilibrium we find:

1. the departure process from this queue is a Poisson process with parameter λ

2. The number in the queue at any time t is independent of the departure process prior to t.

Think of an overstaffed shop.

For chained queues:

1. the arrival process of downstream queues will be poisson
2. the state of downstream queues at time t depend on the departure process of upstream queues **before** t, but is independent of the state of the upstream queues **at** time t

A network of queues ordered from 1 to J such that the customers leaving queue $j \in \{1, 2, \ldots, J\}$ are then fed into queues $j+1, \ldots, J$ or leave the system, then the network has a product form equilibrium distribution:

$$\pi(\mathbf{n}) = \pi(n_1, n_2, \ldots, n_J) = \pi_1(n_1)\pi_2(n_2)\ldots\pi_J(n_J)$$

An open Jackson Network consists of a network of N queues, where at node i of the network:

- Arrivals come from outside of the network at rate λ_i
- The service rate is $\mu_i(n_i)$ when there are n_i customers in that queue,
- a customer upon completing their service will either:
  - move to queue j with probability γ_{ij} or
  - Leave the network with probability $\beta_i = 1 - \sum_j \gamma_{ij}$

An open Jackson Network has the following product form equilibrium distribution (provided it can be normalised)

$$\pi(\mathbf{n}) = \prod_{i=1}^{N} \pi_i(n_i)$$

$$\textit{where} \quad \pi_i(n_i) = \pi_i(0) \prod_{l=1}^{n_i} \frac{y_i}{\mu_i(l)} \quad \textit{is the equilibrium of the ith queue}$$

And y_i is the average arrival rate to queue i, given by the traffic equations:

$$y_i = \lambda_i + \sum_{j=1}^{N} y_j \gamma_{ji}$$

Notes:

- Normalisation depends on whether the constants $\pi_i(0)$ can be found for each $i \in \{1, 2, \ldots, N\}$ such that

$$\sum_{n_i=0}^{\infty} \pi_i(n_i) = 1$$

- We assume that the rate into each queue must be the same as the rate out (for stability purposes). That is, y_i is the arrival rate **and** departure rate from queue i.

invariant measure for the number of customers at queue i if the queue is fed with a Poisson arrival stream of rate y_i:

$$Q_i(n_i) = \prod_{l=1}^{n_i} \frac{y_i}{\mu_i(l)}$$

y_i is the total average arrival rate to queue i but in general is not a Poisson stream.

Equilibrium probabilities π_j can be interpreted as either:

- **Ergodic:** the long term proportion of time the system is in state j

- **Limiting:** the probability that at an arbitrary time point in equilibrium (far away from ICs), the system is in state j.

Let χ be any CTMC, and γ_j be the intensity of an event stream which occurs when χ is in state j. These events may or may not change the state of χ.

Denote $\{\pi_j^{(E)}, j \in S\}$ as the distribution in equilibrium just before particular event points, and $P_j^{(E)}(t)$ be the time-dependent equivalent. We can find general relationships between $\pi_j^{(E)}$ and π_j, and for $P_j^{(E)}(t)$ and $P_j(t)$

We find that:

$$P_j^{(E)}(t) = \frac{\gamma_j P_j(t)}{\sum_{k\in S}\gamma_k P_k(t)} \quad and \quad \pi_j^{(E)} = \frac{\gamma_j \pi_j}{\sum_{k\in S}\gamma_k \pi_k}$$

(PASTA) Poisson arrivals see time averages.
If $\gamma_j = \lambda$ for all $j \in S$, then

$$P_j^{[E]}(t) = P_j(t) \quad and \quad \pi_j^{[E]} = \pi_j$$

For an arbitrary birth/death process the equilibrium distribution as seen by an arrival (that changes the state of the process) can be written as:

$$\pi_j^{(A)} = \pi_0^{(A)} \prod_{l=1}^{j} \frac{\lambda_l}{\mu_l} \quad for \quad j \geq 0$$

(Note this is not the same as the equilibrium $\pi_j = \pi_0 \prod_{l=1}^{j} \frac{\lambda_{l-1}}{\mu_l}$)

1. What is the probability that an arriving customer doesn't have to wait for a server?
   Assuming individual service rate μ and arrival rate λ

$$\begin{aligned} P(W_Q = 0) &= \sum_{j=0}^{N-1} \pi_j^{(A)} \\ &= \sum_{j=0}^{N-1} \pi_j \text{ (using PASTA)} \\ &= \pi_0 \sum_{j=0}^{N-1} \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} \end{aligned}$$

2. What is the distribution of the customer's waiting time?
   This will be Erlang distributed. $Erlang(k+1, N\mu)$

$$\begin{aligned} f_{W_Q}(t) &= \frac{dP(W_Q < t)}{dt} \\ &= \sum_{k=0}^{\infty} P(\text{arrival finds } N+k \text{ in the system}) \times (\text{Erlang}(k+1, N\mu) \text{ density function}) \\ &= \sum_{k=0}^{\infty} \pi_{N+k}^{(A)} N\mu \frac{(N\mu t)^k}{k!} e^{-N\mu t} \\ &= \sum_{k=0}^{\infty} \pi_N \left(\frac{\lambda}{N\mu}\right)^k N\mu \frac{(N\mu t)^k}{k!} e^{-N\mu t} \\ &= \sum_{k=0}^{\infty} \pi_N (\lambda)^k N\mu \frac{(t)^k}{k!} e^{-N\mu t} \\ &= \pi_N N\mu e^{-N\mu t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} \\ &= \pi_N N\mu e^{-N\mu t} e^{\lambda t} \\ &= N\mu e^{(\lambda - N\mu)t} \pi_N \end{aligned}$$

Erlang-C formula

$$f_{W_Q} = C(N, \frac{\lambda}{\mu}) \times (N\mu - \lambda)e^{-(N\mu-\lambda)t}$$

Note that the rhs of this is an exponential pdf with parameter $(N\mu-\lambda)$ - the difference between the maximum service rate and the arrival rate.

3. What is the mean waiting time of an arriving customer?

$$\begin{aligned} E(W_Q) &= \int_0^\infty t f_{W_Q}(t)dt \\ &= \frac{C(N, \frac{\lambda}{\mu})}{N\mu - \lambda} \end{aligned}$$

Alternatively, use $E(W_Q) = C(N, \frac{\lambda}{\mu})\times$ mean of the Exp$(N\mu - \lambda)$

4. What is the conditional waiting time of an arriving customer, given they have to wait?

$$\begin{aligned} P(W_Q > t|W_Q > 0) &= \frac{P(W_Q > t \cap W_Q > 0)}{P(W_Q > 0} \\ &= \frac{P(W_Q > t)}{P(W_Q > 0} \\ &= \frac{C(N, \frac{\lambda}{\mu})e^{-(N\mu-\lambda)t}}{C(N, \frac{\lambda}{\mu})} \\ &= e^{-(N\mu-\lambda)t} \end{aligned}$$

5. Conditional Expectation of the waiting time given you have to wait

$$\begin{aligned} E(W_Q|W_Q > 0) &= (\text{mean of exponential } (N\mu - \lambda) \\ &= \frac{1}{N\mu - \lambda} \end{aligned}$$

Little's Law This theorem gives a relationship between $\bar{L}(\Gamma)$, length of queue, $\bar{W}(\Gamma)$, average waiting time and $\bar{\lambda}(\Gamma)$, the average arrival rate.
In equilibrium,

$$\bar{L}(\Gamma) = \bar{\lambda}(\Gamma)\bar{W}(\Gamma)$$

Pollaczek-Khinchin mean value formulae for a queue Y

$$\begin{aligned} E[q] &= a + \frac{\lambda^2 E[Y^2]}{2(1-a)} \\ E[w] &= \frac{1}{\lambda}\left[\frac{\lambda^2 E[Y^2]}{2(1-a)}\right] \end{aligned}$$

Where q is the queue length, and w is the waiting time, and $a = \frac{\lambda}{\mu}$

Ideal examples of point processes: Poisson and renewal processes.
Point processes are methods of randomly allocating points to intervals of the real line (extendable to hyper-rectangles in n-dimensional Euclidean space)

The stationary Poisson process is defined by:

$$P(N(a_i, b_i] = n_i, i = 1, \ldots, k) = \prod_{i=1}^{k} \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!} e^{-\lambda(b_i - a_i)}$$

Where $N(a_i, b_i]$ is the number of events in the interval $(a_i, b_i]$ for $a_i < b_i \leq a_{i+1}$

1. The number of events in each finite interval $(a_i, b_i]$ is Poisson distributed

2. The number of points in disjoint intervals are independent RVs

3. The process is stationary as the distributions depend only on the lengths $b_i - a_i$ of the intervals, and not the actual end points

As a Poisson RV, then mean and variance are:

$$M(a,b] = \lambda(b-a) = V(a,b]$$

So λ can be interpreted as the mean rate, or mean density of points in the process.
Also, $P(N(0,\tau] = 0) = e^{-\lambda\tau}$, the probability of finding no points in any interval of length τ. This can also be interpreted as:

1. The probability that the random interval extending from the origin to the first point to the right has a length exceeding τ, or

2. The survivor function for the length of this interval

And it shows the interval under consideration has an exponential distribution.
From stationarity, the same result applies to:
The **length** of the interval to the next point immediately to the right/left from any arbitrarily chosen origin. These are called the forward/backward reccurrence times, which are Exp(λ).

Let t_k be the time from the origin $t_0 = 0$ until the kth to the right of the origin. Then the events: $\{t_k \geq x\}$ and $\{N(0,x] \leq k\}$ are equivalent, such that their respective probabilities are the same:

$$P(t_k > x) = P(N(0,x] < k) = \prod_{j=0}^{k-1} \frac{(\lambda x)^j}{j!} e^{-\lambda x}$$

And

$$P(t_k \leq x) = 1 - P(N(0,x] < k) = 1 - \prod_{j=0}^{k-1} \frac{(\lambda x)^j}{j!} e^{-\lambda x}$$

By differentiating this expression we get the density:

$$\begin{aligned} f_k(x) &= \frac{\partial}{\partial x} P(t_k \leq x) \\ &\vdots \\ &= \lambda \frac{(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda x} \end{aligned}$$

This is the density of the Erlang distribution. Which is the sum of the lentghs of the k random intervals $(t_0, t_1], (t_1, t_2], \ldots, (t_{k-1}, t_k]$, which are iid Exp($\lambda$) RVs. (this makes sense since the Erlang is the sum of IID Exp rvs)

Renewal Processes Let $X_1, X_2, \ldots$ be a sequence of iid non-negative random variables with

$$P(X_i \leq x) = F(x) \quad x \geq 0$$

We assume that $F(0) < 1$ and that $E[X_i] = \mu < \infty$. Let

$$\begin{aligned} S_n &= X_1 + X_2 + \ldots + X_n \quad (S_0 = 0) \\ &= \textit{waiting time to the nth event} \\ N(t) &= \sup\{n : S_n \leq t\} \\ &= \textit{number of events before time } t \end{aligned}$$

Then $\{N(t), t > 0\}$ is called the counting process, and S_n is called the waiting time process. The process $\{N(t)\}$ is referred to as a renewal process.
Note that if $F(0) > 0$, then two events can occur simultaneously!

A function $f(t)$ is right continuous at a point τ if for all $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(t) - f(\tau)| < \epsilon \text{ for all } \tau < t < \tau + \delta$$

I.e.

$$\lim_{t \to \tau^+} f(t) = f(\tau)$$

A right-continuous function is a function which is right-continuous at all points.

CDFs are all monotonically non-decreasing

Note there are mirror definitions for left continuous and monotonically non-increasing functions (but we don't care for this course)

If $\alpha(t)$ is a right-continuous, monotonically non-decreasing function on the interval $[a, b]$, define the upper and lower sums with respect to a partition P (where $t_0 = a$ and $t_n = b$) and the function $\alpha(t)$ as

$$\sum_{i=1}^{n} m_i(\alpha(t_i) - \alpha(t_{i-1})) \quad \text{and} \quad \sum_{i=1}^{n} M_i(\alpha(t_i) - \alpha(t_{i-1}))$$

The Riemann-Stieltjes integral of $g(t)$ with respect to $\alpha(t)$ over $[a, b]$ exists if

$$\sup_{\text{all partitions } P} \sum_{i=1}^{n} m_i(\alpha(t_i) - \alpha(t_{i-1})) \quad = \quad \inf_{\text{all partitions } P} \sum_{i=1}^{n} M_i(\alpha(t_i) - \alpha(t_{i-1})) \quad = l$$

In which case we write

$$\int_a^b g(t) d\alpha(t) = l$$

I.e. the Riemann-Stieltjes integral is the Riemann integral of the function

$$g(t) \frac{\partial \alpha(t)}{\partial t}$$

$\alpha(t)$ doesn't have to be differentiable: $\alpha(t)$ may have discontinuities monotonically non-decreasing means its left hand limits must exist. I.e. $\lim_{t \to \tau^-} \alpha(t) = \alpha(\tau^-)$. The function $\alpha(t)$ may be continuous, but not differentiable at τ This means we can integrate piecewise functions!

If $\alpha(t)$ has a discontinuity at a, we want to include it in any integral that starts (or finishes) at a. That is we want to include the mass $(\alpha(a) - \alpha(a^-))$ at a. Thus we interpret the lower limit of the integration to be to the left of a, and the upper limit to the write. So for integration we **could** write:

$$\int_{a^-}^{b^+} g(t) d\alpha(t) = l$$

If $g(t)$ and $\alpha(t)$ have a discontinuity at the same point τ, the upper and lower sums will differ by at least

$$[g(\tau) - g(\tau^-)[\alpha(\tau) - \alpha(\tau^-)]$$

This can only really be resolved by resorting to more painful measure theory. To solve this we regard the correct contribution to the integral as

$$g(\tau)(\alpha(\tau) - \alpha(\tau^-))$$

We use the Laplace-Stieltjes transforms a lot in renewal theory.

Let X be a non-negative RV with distribution function $F(x)$. Then the Laplace-Stieltjes transform of X is given by:

$$\hat{F}(s) = \int_0^\infty e^{-sx} dF(x)$$

Noting that $\hat{F}(s) = E(e^{-sX})$

If X and Y are independent RVs, then the RV $Z = X + Y$ is known as the convolution of X and Y.
Z has the distribution function:

$$F_Z(z) = \int_0^z F_X(z - y) dF_Y(y)$$

Convolution Theorem: For independent RVs X_i, $i = 1, \dots n$, the RV $Z \sum_{i=1}^{n} X_i$ has the Laplace-Stieltjes Transform

$$\hat{F}_Z(s) = \prod_{i=1}^{n} \hat{F}_{X_i}(s)$$

The Renewal function (mean value function) $M(t)$ is defined as

$$M(t) = E[N(t)] = \sum_{n=0}^{\infty} nP_n(t)$$

$M(t)$ is the expected number of events to occur by time t.
If $F(0) < 1$, then $M(t) < \infty$ for $t > 0$.
Letting $\hat{M}(s) = \int_0^{\infty} e^{-st} dM(t)$ we get

$$\hat{M}(s) = \sum_{n=1}^{\infty} (\hat{F}(s))^n$$

And since $\hat{F}(s) < 1$ because

$$\hat{F}(s) = \int_0^{\infty} e^{-st} dF(t) < \int_0^{\infty} 1 dF(t) = 1$$

And $F(0) < 1$ means that there must be some contribution to both of the above integrals for a positive value of t, for which $e^{-st} < 1$.
So for $s > 0$, we end up getting

$$\hat{M}(s) = \frac{\hat{F}(s)}{1 - \hat{F}(s)}$$

The Renewal function $M(t)$ satisfies the *renewal equation*

$$M(t) = F(t) + \int_0^t M(t - x) dF(x)$$

Note that renewal theory is useful due to the ability to condition on the time of the first renewal.
The only solution to the renewal equation, which is bounded on finite intervals, is given by

$$M(t) = \sum_{n=1}^{\infty} F_n(t)$$

$$\hat{M}(s) = \frac{\hat{F}(s)}{1 - \hat{F}(s)}$$

And consequently

$$\hat{F}(s) = \frac{\hat{M}(s)}{1 + \hat{M}(s)}$$

The generalised renewal equation for $H(t)$ (the expected number of events by time t):

$$H(t) = G(t) + \int_0^t H(t - y) dF(y)$$

Where $G(t)$ is the distribution of the first lifetime, and $F(t)$ is the distribution of each of the subsequent lifetimes. In the normal renewal equations, $F(t) = G(t), \forall t$.
The solution to the generalised renewal equation is:

$$H(t) = G(t) + \int_0^t G(t - y) dM(y)$$

Where $M(t)$ is the solution to

$$M(t) = F(t) + \int_0^t M(t - x) dF(x)$$

$$H'(s) = \mathcal{L}^{-1}(\frac{\hat{G}(s)}{1-\hat{F}(s)})$$

and then

$$H(t) = \int_0^t H'(s)ds$$

For any renewal process, the expected time between the start of the process until the first event after time t

$$E(S_{N(t)+1}) = \mu(M(t)+1)$$

Let $N(t) : t \geq 0$ be a counting process. Then

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mu}$$

with probability 1, where μ is the mean time between events.

Renewal Theorem: Let $F(t)$ be the distribution function of a positive RV with mean $\mu < \infty$, and assume that $F(t)$ is not *lattice* (does not have all of its points of increase at multiples of some δ.)
Suppose that $H(t)$ is a solution of the generalised renewal equation

$$H(t) = G(t) + \int_0^t H(t-y)dF(y)$$

Where $G(t)$ is integrable
Then

$$\lim_{t\to\infty} H(t) = \frac{1}{\mu}\int_0^\infty G(t)dt$$

If F *is* lattice, then this is valid if $t = n\delta$ for all $n \in N$.

Blackwell's Renewal Theorem Let F be the distribution function of a positive RV with mean $\mu < \infty$ which is not lattice.

$$\lim_{t\to\infty}[M(t) - M(t-h)] = \frac{h}{\mu}, \quad for\ h > 0$$

Elementary Renewal theorem: If $F(t)$ is not lattice

$$\lim_{t\to\infty} \frac{M(t)}{t} = \frac{1}{\mu}$$

3 Time Series TS

3.1 Formulae:

How to express an ARMA process as a GLP

3.1.1 Spectra:

General formula:

$$f_y(w) = |a(w)|^2 f_v(w)$$

For moving average (MA) of form:

$$Y_t = \theta(B)Z_t$$

The spectrum is given by:

$$f_y(w) = |\theta(e^{iw})|^2\sigma^2$$

For autoregressive (AR) of form:

$$\phi(B)Y_t = Z_t$$

The spectrum is given by:

$$f_y(w) = \frac{\sigma^2}{|\phi(e^{iw})|^2}$$

For autoregressive-moving average (ARMA) of form:

$$\phi(B)Y_t = \theta(B)Z_t$$

The spectrum is given by:

$$f_y(w) = \sigma^2\frac{|\theta(e^{iw})|^2}{|\phi(e^{iw})|^2}$$

Note that the forms of ϕ and θ used in the formulas yield:

$$|\phi(e^{iw})|^2 = \left(1 - \sum_{j=1}^{p}\alpha_j\cos(wj)\right)^2 + \left(\sum_{j=1}^{p}\alpha_j\sin(wj)\right)^2$$

and

$$|\theta(e^{iw})|^2 = \left(1 + \sum_{j=1}^{q}\beta_j\cos(wj)\right)^2 + \left(\sum_{j=1}^{q}\beta_j\sin(wj)\right)^2$$

3.2 Proofs/Tricks:

3.3 Notes:

The Exam **WILL** be:

1. General definitions (stationarity, etc.)
2. The AR(1) process
3. Moving average processes
4. Autoregressive processes
5. Practical Question - data analysis skills and interpretation

Where the practical question is 20 marks and the rest add up to 50.
Diggle model:
Start with a time plot

1. Does it appear to be stationary?

- No? Difference the data and restart at 1
- Yes? continue

2. Does the correlogram of the data decay to zero?
    - No? Difference the data and restart at 1 again
    - Yes? Continue
3. Is there a sharp cut-off in the correlogram?
    - Yes? It is a **moving average** model
    - No? continue
4. Is there a sharp cut-off in the partial correlogram?
    - Yes? It is an **autoregressive model**
    - No? It is an **ARMA** model

If we see exponential decay in the ACF/PACF and a distinct cut off in the PACF/ACF (other one), then we can model with just the AR / MA model respectively.