ESTIMATION AND GOODNESS OF FIT

Method of Moments: $m_j = \frac{1}{n}\sum_{i=1}^{n} x_i^j \quad j = 1, 2 \ldots r$

with x_i = the i^{th} value in the sample & n = the sample size

Likelihood Function $L(\theta)$:

$L(\theta) = \prod_{i=1}^{n} P(X_i = x_i \mid \theta)$ (discrete)

$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$ (continuous)

Log-likelihood Function $L(\theta)$:

$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log P(X_i = x_i \mid \theta)$ (discrete)

$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i \mid \theta)$ (continuous)

Exponential Distribution: $F(x) = 1 - e^{-\lambda x},\ x > 0$ $\quad f(x) = \lambda e^{-\lambda x},\ x > 0$

Gamma Distribution: $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},\ x > 0$ $\quad E(X) = \frac{\alpha}{\lambda} \quad Var(X) = \frac{\alpha}{\lambda^2}$

Lognormal Distribution: $\log(X) \sim N(\mu, \sigma^2)$ then $X \sim \log N(\mu, \sigma^2)$ $\quad$ MLEs of μ and σ^2 are $\bar{y}$ and s_y^2

Pareto Distribution: $F(x) = 1 - \left(\frac{\lambda}{\lambda + x}\right)^\alpha,\ x > 0$ $\quad f(x) = \frac{\alpha\lambda^\alpha}{(\lambda + x)^{\alpha+1}},\ x > 0$

Burr Distribution: $F(x) = 1 - \frac{\lambda^\alpha}{(\lambda + x^\gamma)^\alpha},\ x > 0$

Three-Parameter Pareto Distribution: $f_X(x) = \frac{\Gamma(\alpha + k)\lambda^\alpha}{\Gamma(\alpha)\Gamma(k)} \frac{x^{k-1}}{(\lambda + x)^{\alpha+k}},\ x > 0$

Weibull Distribution: $F(x) = 1 - e^{-cx^\gamma},\ x > 0$ $\quad f(x) = c\gamma x^{\gamma-1} e^{-cx^\gamma},\ x > 0$

REINSURANCE

Excess of Loss Reinsurance: Payment of Insurer $(Y) = \begin{cases} X & \text{if } X \le M \\ M & \text{if } X > M \end{cases}$

$E(Y) = \int_0^M x f(x)dx + M\,\mathrm{P}(X > M)$

MGF of Y: $M_Y(t) = E\left(e^{tY}\right) = \int_0^M e^{tx} f(x)dx + e^{tM} P(X > M)$

Payment of Reinsurer $(Z) = X - Y$

$E(Z) = \int_M^\infty (x - M)f(x)dx$

Effect of Inflation: Payment of Insurer $(Y) = \begin{cases} kX & \text{if } kX \le M \\ M & \text{if } kX > M \end{cases}$

$E(Y) = \int_0^{M/k} kx f(x)dx + M\,\mathrm{P}(X > M/k)$

Likelihood Function: $L(\underline{\theta}) = \prod_1^n f(x_i; \underline{\theta}) \times [(1 - F(M; \underline{\theta}))]^m$

Reinsurer Claim: $P(W < w) = \int_M^{w+M} \frac{f(x)}{1 - F(M)}dx = \frac{F(w + M) - F(M)}{1 - F(M)}$ $\quad g(w) = \frac{f(w + M)}{1 - F(M)}$

Proportional Reinsurance: Insurer: $Y = \alpha X,\ 0 < \alpha < 1$ $\quad$ Reinsurer: $Z = (1 - \alpha)X$

Mean: $E(Y) = \alpha E(X)$ $\quad E(Z) = (1 - \alpha)E(X)$

Variance: $\mathrm{Var}(Y) = \alpha^2\,\mathrm{Var}(X)$ $\quad \mathrm{Var}(Z) = (1 - \alpha)^2\,\mathrm{Var}(X)$

Policy Excess: $Y = \begin{cases} 0 & \text{if } X \le L \\ X - L & \text{if } X > L \end{cases}$

RISK MODELS

Collective Risk Model:

$S = X_1 + X_2 + \cdots + X_N = \sum_{i=1}^{N} X_i$

$P(S \le x) = \sum_{n=0}^{\infty} P(N=n)P(S \le x \mid N=n)$

$E[S] = E[E(S \mid N)] = E[N]m_1 = E(N)E(X)$

$Var[S] = E[Var[S \mid N]] + Var[E[S \mid N]] = E[N]\left(m_2 - m_1^2\right) + Var[N]m_1^2$
$= E(N)V(X) + V(N)[E(X)]^2$

$M_S(t) = E[E[\exp\{tS\} \mid N]]$ $\quad E[\exp\{tS\} \mid N] = [M_X(t)]^N$

Compound Poisson:

$E[S] = \lambda m_1$ $\quad Var[S] = \lambda m_2$

$M_S(t) = \exp\{\lambda(M_X(t) - 1)\}$ $\quad \text{skew}[S] = \lambda m_3$

$M(t) = \frac{1}{\Lambda}\sum_{i=1}^{n} \lambda_i \int_0^{\infty} \exp\{tx\} f_i(x)dx = \frac{1}{\Lambda}\sum_{i=1}^{n} \lambda_i M_i(t)$

$M_A(t) = \exp\{\Lambda(M(t) - 1)\}$

Compound Binomial:

$E[S] = npm_1$ $\quad Var[S] = npm_2 - np^2m_1^2$

$M_S(t) = (pM_X(t) + 1 - p)^n$ $\quad \text{skew}[S] = npm_3 - 3np^2m_2m_1 + 2np^3m_1^3$

Compound Negative Binomial:

$E[S] = \frac{kq}{p}m_1$ $\quad Var[S] = \frac{kq}{p}m_2 + \frac{kq^2}{p^2}m_1^2$

$M_S(t) = p^k / (1 - qM_X(t))^k$ $\quad \text{skew}[S] = \frac{3kq^2m_1m_2}{p^2} + \frac{2kq^3m_1^3}{p^3} + \frac{kqm_3}{p}$

Individual Excess of Loss: $S_I = Y_1 + Y_2 + \ldots + Y_N$ $\quad S_R = Z_1 + Z_2 + \ldots + Z_N$

Aggregate Excess of Loss: $S_I = \min(S, M)$ $\quad S_R = \max(0, S - M)$

Individual Risk Model: $S = Y_1 + Y_2 + \ldots + Y_n$

Assumptions:

The number of claims from the j-th risk, N_j, is either 0 or 1

The probability of a claim from the j-th risk is q_j

$E[Y_j] = q_j\mu_j$ $\quad \text{Var}[Y_j] = q_j\sigma_j^2 + q_j(1-q_j)\mu_j^2$

$E[S] = \sum_{j=1}^{n} E[Y_j] = \sum_{j=1}^{n} q_j\mu_j$ $\quad \text{Var}[S] = \sum_{j=1}^{n} \text{Var}[Y_j] = \sum_{j=1}^{n}\left(q_j\sigma_j^2 + q_j(1-q_j)\mu_j^2\right)$

COPULAS: MARGINAL AND JOINT DISTRIBUTIONS

Bivariate Case: $C_{XY}[F_X(x), F_Y(y)] = P(X \le x, Y \le y) = F_{X,Y}(x,y)$

Multivariate Case: $C[u_1, u_2, \ldots, u_d] = F_{X_1,X_2,\ldots,X_d}(x_1, x_2, \ldots, x_d)$ where $u_i = F_{X_i}(x_i)$

Product Copula: $C[u,v] = uv$

Co-monotonic Copula: $C[u,v] = \min(u,v)$

Counter-monotonic Copula: $C[u,v] = \max(u + v - 1, 0)$

Gumbel Copula: $C(u,v) = \exp\left\{-\left((-\ln u)^{\alpha} + (-\ln v)^{\alpha}\right)^{\frac{1}{\alpha}}\right\}, \alpha \ge 1$

Frank Copula: $C(u,v) = -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{(e^{-\alpha}-1)}\right), \alpha \ne 0$

Clayton Copula: $C(u,v) = \left(\max\left(u^{-\alpha} + v^{-\alpha} - 1, 0\right)\right)^{-\frac{1}{\alpha}}, \ \alpha \ge -1, \alpha \ne 0$

$C(u,v) = \left(u^{-\alpha} + v^{-\alpha} - 1\right)^{-\frac{1}{\alpha}}, \ \alpha > 0$

Sklar's Theorem: Let F be a joint distribution function with marginal cumulative distribution functions $F_1, \ldots, F_d$.

Then there exists a copula C such that for all $\boldsymbol{x_1}, \ldots, \boldsymbol{x_d} \in [-\infty, \infty]$:
$F(x_1, \ldots, x_d) = C[F_1(x_1), \ldots, F_d(x_d)]$

In the case of variables that have a continuous distribution, the copula C is unique.

If C is a copula and $F_1, \ldots, F_d$ are univariate cumulative distribution functions, then the function F defined above is a joint cumulative distribution function with marginal cumulative distribution functions $F_1, \ldots, F_d$.

Gaussian Copula: $C(u,v) = \Phi_\rho\left[\Phi^{-1}(u), \Phi^{-1}(v)\right]$

Student's t Copula: $C(u,v) = t_{\gamma,\rho}\left[t_\gamma^{-1}(u), t_\gamma^{-1}(v)\right]$

Pseudo-Inverse Function: $\psi^{[-1]}(x) = \begin{cases} \psi^{-1}(x) & \text{if } 0 \le x \le \psi(0) \\ 0 & \text{if } \psi(0) < x \le \infty \end{cases}$

Generator Formulation of Archimedean Copula

Copulas in the Archimedean family:

$$C(u_1, \ldots, u_d) = \psi^{[-1]}\left(\sum_{i=1}^{d} \psi(u_i)\right)$$

The generator function $\psi : [0,1] \to [0,\infty]$ must be a continuous, strictly decreasing, convex function with $\psi(1) = 0$.

Bivariate Case: $C(u,v) = \psi^{[-1]}(\psi(u) + \psi(v))$

Gumbel(-Hougaard) Copula: $\psi(t) = (-\ln t)^\alpha$ where $1 \le \alpha < \infty$

Clayton Copula: $\psi(t) = -\ln\left(\dfrac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}\right)$ where $-\infty < \alpha < \infty, \alpha \neq 0$

Frank Copula: $\psi(t) = \dfrac{1}{\alpha}(t^{-\alpha} - 1)$ where $-1 \le \alpha < \infty, \alpha \neq 0$

COPULAS: CONCORDANCE AND TAIL DEPENDENCE

Invariance Property of a Measure of Concordance:

Two commonly used measures of concordance: Spearman's ρ and Kendall's τ

The measure of concordance does not change if we apply the same monotone function to the value of each variable

Coefficient of Lower Tail Dependence: $\lambda_L = \lim_{u \to 0^+} P\left(X \le F_X^{-1}(u) \mid Y \le F_Y^{-1}(u)\right) = \lim_{u \to 0^+} \dfrac{C(u,u)}{u}$

Coefficient of Upper Tail Dependence: $\lambda_U = \lim_{u \to 1^-} P\left(X > F_X^{-1}(u) \mid Y > F_Y^{-1}(u)\right) = \lim_{u \to 0^+} \dfrac{\bar{C}(u,u)}{u}$

Tail Dependence for Frank Copula: Zero dependence in both tails

Tail Dependence for Gaussian Copula: Zero dependence in both tails

Tail Dependence for Student's t Copula: Equal positive dependence in both tails

Tail Dependence for Gumbel Copula: Zero lower tail dependence but upper tail dependence of $2 - 2^{1/\alpha}$

Tail Dependence for Clayton Copula: Zero upper tail dependence but lower tail dependence of $2^{-1/\alpha}$

EXTREME VALUE THEORY

Block Maximum: $X_M = \max(X_1, X_2, \ldots, X_n)$, the maximum value in a set of n values

Generalised Extreme Value: $H(x) = \begin{cases} \exp\left(-\left(1 + \dfrac{\gamma(x-\alpha)}{\beta}\right)^{-1/\gamma}\right) & \gamma \neq 0 \\ \exp\left(-\exp\left(-\dfrac{(x-\alpha)}{\beta}\right)\right) & \gamma = 0 \end{cases}$

	GEV distributions (for the maximum value) corresponding to common loss distributions		
Type Shape parameter	**WEIBULL** $\gamma < 0$	**GUMBEL** $\gamma = 0$	**FRÉCHET** $\gamma > 0$
Underlying Distribution	Beta Uniform Triangular	Chi-square Exponential Gamma Log-normal Normal Weibull	Burr F Log-gamma Pareto t
Range of values permitted	$x < \alpha - \dfrac{\beta}{\gamma}$	$-\infty < x < \infty$	$x > \alpha - \dfrac{\beta}{\gamma}$

Threshold Exceedance: let X be a random variable with cumulative distribution function F then the excess over the threshold u is $X - u \mid X > u$.

Generalised Pareto Distribution:

$$G(x) = \begin{cases} 1 - \left(1 + \dfrac{x}{\gamma\beta}\right)^{-\gamma} & \gamma \neq 0 \\ 1 - \exp\left(-\dfrac{x}{\beta}\right) & \gamma = 0 \end{cases}$$

Existence of Moments: The more moments exist, the less the tail weight.

Limiting Density Ratios:

$$\text{For } \lim_{x\to\infty} \frac{f_1(x)}{f_2(x)} = c, \begin{cases} \text{if } c = 0, \text{ then } f_1(x) \text{ has thicker tail} \\ \text{if } c = \infty, \text{ then } f_0(x) \text{ has thicker tail} \\ \text{If } 0 < c < \infty, f_1(x) \text{ and } f_2(x) \text{ have similar thickness of tails} \end{cases}$$

Hazard Rate: If hazard rate function is increasing then it has a light tail, it has thick tail if it is decreasing

Mean Residual Life:

$$e(x) = \frac{\int_x^\infty (y - x) f(y)dy}{\int_x^\infty f(y)dy} = \frac{\int_x^\infty \{1 - F(y)\}dy}{1 - F(x)}$$

STOCHASTIC PROCESSES

Strict Stationarity: Statistical properties of the process remain unchanged as time elapses

Weak Stationarity:

Mean of the process: $m(t) = E[X_t]$ is constant

Covariance of the process: $C(s,t) = E[(X_s - m(s))(X_t - m(t))]$, depends only on the time difference $t - s$.

Markov Property: X_t has the Markov property if $P[X_t \in A \mid X_{s_1} = x_1, X_{s_2} = x_2, \ldots, X_{s_n} = x_n, X_s = x] = P[X_t \in A \mid X_s = x]$ for all times $s_1 < s_2 < \ldots < s_n < s < t$, all states $x_1, x_2, \ldots, x_n$, x in S and all subsets A of S

White Noise Processes: It has a mean of zero at all times, it is stationary

Mean: $m(t) = E[X_t] = 0$ for all values of t

Covariance: $C(s,t) = E[(X_s - m(s))(X_t - m(t))]$ is zero for $s \neq t$

General Random Walk: $X_n = \sum_{j=1}^{n} Y_j$ with initial condition $X_0 = 0$, it is non-stationary

Simple Random Walk: The steps Y_j of the walk take only the values $+1$ and -1

Poisson Process with rate λ is a continuous-time integer-valued process $N_t, t \geq 0$

Properties: (i) $N_0 = 0$

(ii) N_t has independent increments

(iii) N_t has Poisson distributed stationary increments

$$P\left[N_t - N_s = n\right] = \frac{[\lambda(t-s)]^n e^{-\lambda(t-s)}}{n!}, \quad s < t, n = 0, 1, \ldots$$

Compound Poisson Process: $X_t = \sum_{j=1}^{N_t} Y_j,\ t \geq 0$ with a Poisson process $N_t,\ t \geq 0$ and a sequence of IID random variables $Y_j,\ j \geq 1$

Probability of Ruin: $\Psi(u) = P\left[u + ct - X_t < 0 \text{ for some } t > 0\right]$

Purely Indeterministic: the knowledge of the values of $X_1, \ldots, X_n$ is progressively less useful at predicting the value of X_N as $N \to \infty$.

TIME SERIES MODELS

Autocovariance Function: $\gamma_k \equiv \operatorname{Cov}\left(X_t, X_{t+k}\right) = E\left(X_t X_{t+k}\right) - E\left(X_t\right) E\left(X_{t+k}\right)$

Autocorrelation Function: $\rho_k = \operatorname{Corr}\left(X_t, X_{t+k}\right) = \frac{\gamma_k}{\gamma_0}$

Partial Autocorrelation Function: ϕ_k, The conditional correlation of X_{t+k} with X_t given $X_{t+1}, \ldots X_{t+k-1}$

$\phi_1 = \rho_1,\ \phi_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$, ϕ_k depends on only $\rho_1, \rho_2, , , \rho_k$

Autoregressive Process: $X_t = \mu + \alpha_1\left(X_{t-1} - \mu\right) + \alpha_2\left(X_{t-2} - \mu\right) + \ldots + \alpha_p\left(X_{t-p} - \mu\right) + e_t$

Moving average Process: $X_t = \mu + e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q}$

ARMA Process: $X_t = \mu + \alpha_1\left(X_{t-1} - \mu\right) + \ldots + \alpha_p\left(X_{t-p} - \mu\right) + e_t + \beta_1 e_{t-1} + \ldots + \beta_q e_{t-q}$

Backwards Shift Operator B: $BX_t = X_{t-1}$

Difference Operator, ∇ $\nabla = 1 - B \quad \rightarrow \quad \nabla X_t = X_t - X_{t-1}$

First-order Autoregressive AR(1):

$$X_t = \mu + \alpha\left(X_{t-1} - \mu\right) + e_t = \mu + \alpha^t\left(X_0 - \mu\right) + \sum_{j=0}^{t-1} \alpha^j e_{t-j}$$

Mean: $\mu_t = \mu + \alpha^t\left(\mu_0 - \mu\right)$

Variance: $\operatorname{Var}\left(X_t\right) = \sigma^2 \frac{1 - \alpha^{2t}}{1 - \alpha^2} + \alpha^{2t} \operatorname{Var}\left(X_0\right)$

Autocovariance: $\gamma_k = \operatorname{Cov}\left(X_t, X_{t+k}\right) = \sum_{j=0}^{\infty} \sigma^2 \alpha^{2j+k} = \alpha^k \gamma_0$

PACF: $\phi_1 = \rho_1 = \alpha,\ \phi_k = 0$ for all $k > 1$

ACF: ρ_k decreases geometrically towards 0

Force of Inflation: $r_t = \nabla \log\left(Q_t\right) = \mu + \alpha\left(r_{t-1} - \mu\right) + e_t$

Autoregressive model $AR(p)$: $X_t = \mu + \alpha_1\left(X_{t-1} - \mu\right) + \alpha_2\left(X_{t-2} - \mu\right) + \cdots + \alpha_p\left(X_{t-p} - \mu\right) + e_t$

$AR(p)$, Backward Shift Formulation:

$$\left(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p\right)\left(X_t - \mu\right) = e_t$$

Characteristic Polynomial: $1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p = 0$

Stationary Criterion: An $AR(p)$ process is stationary if the roots of $\alpha_p B = 0$ greater than 1 in absolute value.

Yule-Walker Equation: $\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \cdots + \alpha_p \gamma_{k-p} + \sigma^2 \mathbf{1}_{\{k=0\}}$ where $\mathbf{1}_{\{k=0\}} = 1$ if $k = 0$, 0 otherwise

First-order Moving Average $MA(1)$:

$$X_t = \mu + e_t + \beta e_{t-1}$$

$MA(1)$, Backward Shift Formulation:

$$X - \mu = (1 + \beta B)e$$

Mean: $\mu_t = \mu$

Variance: $\operatorname{Var}(e_t + \beta e_{t-1}) = \gamma_0 = (1 + \beta^2)\sigma^2$

Autocovariance: $\operatorname{Cov}(e_t + \beta e_{t-1}, e_{t-1} + \beta e_{t-2}) = \gamma_1 = \beta\sigma^2$
$\gamma_k = 0$ for $k > 1$

PACF: $\phi_k = (-1)^{k+1} \dfrac{(1-\beta^2)\beta^k}{1 - \beta^{2(k+1)}}$

ACF: $\rho_0 = 1,\ \rho_1 = \dfrac{\beta}{1+\beta^2},\ \rho_k = 0$ for $k > 1$

Moving Average Model $MA(q)$: $X_t - \mu = (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^p)\, e$

Autocovariance Function: $\gamma_k = \sum_{i=0}^{q} \sum_{j=0}^{q} \beta_i \beta_j E(e_{t-i} e_{t-j-k}) = \sigma^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k}$

Invertibility Criterion: An $MA(q)$ process is invertible if roots of $\beta_q B = 0$ greater than 1 in absolute value.

ARMA(p,q) Process: $(1 - \alpha_1 B - \cdots - \alpha_p B^p)(X_t - \mu) = (1 + \beta_1 B + \cdots + \beta_q B^q)\, e_t$

ARMA(1,1) Process: $X_t = \alpha X_{t-1} + e_t + \beta e_{t-1}$

$$\rho_1 = \frac{(1+\alpha\beta)(\alpha+\beta)}{(1+\beta^2+2\alpha\beta)} \qquad \rho_k = \alpha^{k-1}\rho_1,\ k = 2, 3, \ldots$$

$$\gamma_0 = \frac{1 + 2\alpha\beta + \beta^2}{1-\alpha^2}\sigma^2, \qquad \gamma_1 = \frac{(\alpha+\beta)(1+\alpha\beta)}{1-\alpha^2}\sigma^2, \qquad \gamma_k = \alpha^{k-1}\gamma_1$$

ARIMA(p,d,q) Process: $(1 - \alpha_1 B - \cdots - \alpha_p B^p)(1-B)^d (X_t - \mu) = (1 + \beta_1 B + \cdots + \beta_q B^q)\, e_t$

Markov Property: MA(q) processes are not Markov

AR(p) processes are Markov only when $p = 1$

ARMA(p,q) processes are Markov only when $p = 1$ and $q = 0$

TIME SERIES MODELS 2

Seasonal Differencing Model: $x_t = \mu + \theta_t + y_t$
where θ is a periodic function with period 12 and y is a stationary series

Method of Moving Averages: $y_t = \dfrac{1}{2h}\left(\dfrac{1}{2}x_{t-h} + x_{t-h+1} + \cdots + x_{t-1} + x_t + \cdots + x_{t+h-1} + \dfrac{1}{2}x_{t+h}\right)$

Method of Seasonal Means: $\hat{\theta}_t = \dfrac{1}{n}\left(x_t + x_{t+12*1} + \cdots + x_{t+12*(n-1)}\right) - \hat{\mu}$

Stationary Time Series:

Sample mean: $\hat{\mu} = \dfrac{1}{n}\sum_{t=1}^{n} x_t$

Sample autocovariance function: $\hat{\gamma}_k = \dfrac{1}{n}\sum_{t=k+1}^{n} (x_t - \hat{\mu})(x_{t-k} - \hat{\mu})$

Sample autocorrelation function: $\hat{\rho}_k = \dfrac{\hat{\gamma}_k}{\hat{\gamma}_0}$

Sample partial autocorrelation function $\hat{\phi}_k$: Use the formula involving the ratio of determinants with the ρ_k replaced by their estimates $\hat{\rho}_k$

'Portmanteau' Test: $n(n+2)\sum_{k=1}^{m} \dfrac{r_k^2}{n-k} \sim \chi_m^2$ for each m

Identification of $MA(q)$: $\hat{\rho}_k$ is close to 0 for all $k > q$ and $\tilde{\rho}_k$ is normally distributed with mean 0, variance $n^{-1}\left(1+2\sum_{k=1}^{q}\rho_k^2\right)$

Identification of $AR(p)$: $\hat{\phi}_k$ is zero for $k > p$ and asymptotic variance of $\tilde{\phi}_k$ is $1/n$ for each $k > p$

Akaike Information Criterion: $\text{AIC} = -2 \times \log L_M + 2 \times (\text{number of parameters})$

Box-Jenkins k-step ahead forecast:

1. replacing all (unknown) parameters by their estimated values
2. Replace the random variables $X_1, \dots, X_n$ with their observed values $x_1, \dots, x_n$
3. Replace the random variables $X_{n+1}, \dots, X_{n+k-1}$ with their forecast values $\hat{x}_n, \dots, \hat{x}_n(k-1)$
4. Replace the innovations $e_1, \dots, e_n$ with the residuals $\hat{e}_1, \dots, \hat{e}_n$
4. Replace the random variables $e_{n+1}, \dots, e_{n+k-1}$ with their expectations, which is 0

Exponential Smoothing: $\hat{x}_n(1) = (1-\alpha)\hat{x}_{n-1}(1) + \alpha x_n = \hat{x}_{n-1}(1) + \alpha\left(x_n - \hat{x}_{n-1}(1)\right)$

Vector autoregressive Process: VAR(p) is a sequence of m-component random vectors $\{\mathbf{X}_1, \mathbf{X}_2, \dots\}$ satisfying $\mathbf{X}_t = \mu + \sum_{j=1}^{p} A_j\left(\mathbf{X}_{t-j} - \mu\right) + \mathbf{e}_t$

Relation between Interest Rates and Tendency to Invest: $\begin{cases} i_t - \mu_i = \alpha_{11}\left(i_{t-1} - \mu_i\right) + e_t^{(i)} \\ I_t - \mu_I = \alpha_{21}\left(i_{t-1} - \mu_i\right) + \alpha_{22}\left(I_{t-1} - \mu_I\right) + e_t^{(I)} \end{cases}$

Cointegrated Time Series: X and Y are cointegrated: (i) X and Y are $I(1)$ random processes.

(ii) There exists a non-zero vector (α, β) such that $\alpha X + \beta Y$ is stationary.

Bilinear Model: $X_n - \alpha\left(X_{n-1} - \mu\right) = \mu + e_n + \beta e_{n-1} + b\left(X_{n-1} - \mu\right)e_{n-1}$

Threshold Autoregressive Model: $X_n = \mu + \begin{cases} \alpha_1\left(X_{n-1} - \mu\right) + e_n, & \text{if } X_{n-1} \le d, \\ \alpha_2\left(X_{n-1} - \mu\right) + e_n, & \text{if } X_{n-1} > d. \end{cases}$

Random Coefficient Autoregressive Model: $X_t = \mu + \alpha_t\left(X_{t-1} - \mu\right) + e_t$

Autoregressive Models with Conditional Heteroscedasticity:

$$\text{ARCH}(p):\ X_t = \mu + e_t\sqrt{\alpha_0 + \sum_{k=1}^{p}\alpha_k\left(X_{t-k} - \mu\right)^2}$$

$$\text{ARCH}(1):\ X_t = \mu + e_t\sqrt{\alpha_0 + \alpha_1\left(X_{t-1} - \mu\right)^2}$$

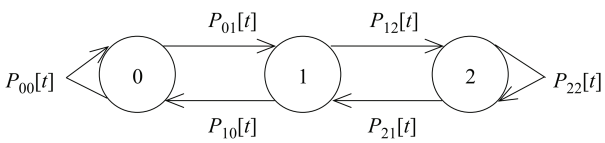
MARKOV CHAINS

Chapman-Kolmogorov Equations: $p_{ij}^{(m,n)} = \sum_{k \in S} p_{ik}^{(m,\ell)} p_{kj}^{(\ell,n)}$ for all states i, j in S and all integer times $m < \ell < n$

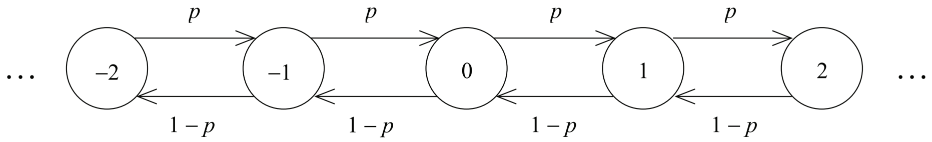
$$P\left[X_n = j \mid X_m = i\right] = \sum_{k \in S} P\left[X_n = j \mid X_\ell = k\right] P\left[X_\ell = k \mid X_m = i\right]$$

ℓ-step Transition Probability: $P\left[X_{m+l} = j \mid X_m = i\right] = p_{ij}^{(\ell)}$ $\quad \sum_{j \in S} p_{ij}^{(\ell)} = 1$ for all i

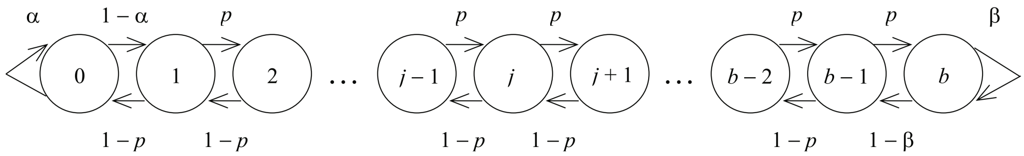
Transition Graph of Time-Inhomogeneous NCD Simple Model:



Transition Graph of Simple Random Walk:



Transition Graph of Simple Random Walk with Boundary:



THE LONG-TERM PROBABILITY DISTRIBUTION OF A MARKOV CHAIN

$\pi_j, j \in S$ is a **stationary probability distribution** for a Markov chain with transition matrix P if the following conditions hold for all j in S:

Conditions: $\pi_j = \sum_{i \in S} \pi_i p_{ij}, \quad \pi_j \geq 0, \quad \sum_{j \in S} \pi_j = 1$

Irreducible Markov Chain: Any state j can be reached from any other state i

Periodic/Aperiodic State: A chain where re-entry to states can only occur every d periods for $d > 1$ is periodic and it is aperiodic when $d = 1$

Best Estimate of Transition Probability for Markov Chain:

$\hat{p}_{ij} = \dfrac{n_{ij}}{n_i}$ with n_i as the number of times $t(1 \leq t \leq N-1)$ such that $x_t = i$,

n_{ij} as the number of times $t(1 \leq t \leq N-1)$ such that $x_t = i$ and $x_{t+1} = j$

Test Statistic for Chi-Square GoF of Markov Chain Estimation: $X^2 = \sum_i \sum_j \sum_k \dfrac{(n_{ijk} - n_{ij}\hat{p}_{jk})^2}{n_{ij}\hat{p}_{jk}}$

Summary:

A Markov chain with a finite state space has at least one stationary distribution

An irreducible Markov chain with a finite state space has a unique stationary distribution

Let $p_{ij}^{(n)}$ be the n-step transition probability of an irreducible aperiodic Markov chain on a finite state space,

Then for every i and j: $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$ where π is the stationary probability distribution

MARKOV JUMP PROCESSES

Transition Probability: $p_{ij}(t) = \Pr\left[X_t = j \mid X_0 = i\right]$

Chapman-Kolmogorov Equation: $p_{ij}(t+s) = \sum_{k \in S} p_{ik}(s) p_{kj}(t) \quad \forall s, t > 0$

Kolmogorov Forward Equation: $\dfrac{d}{dt} p_{ij}(t) = \sum_{k \in S} p_{ik}(t)\mu_{kj}$, all i, j

$\dfrac{d}{dt} P(t) = P(t)A$ where A is the matrix with entries μ_{kj}

Kolmogorov Backward Equation: $\dfrac{d}{dt} P(t) = AP(t)$

Memoryless Property: $\Pr[T > t + u \mid T > t] = \Pr[T > u]$

KFE for Two-State Model: $\dfrac{\partial}{\partial t} p_{AA}(s,t) = -\mu(t) p_{AA}(s,t)$

Survival Probability: $p_{AA}(s,t) = e^{-\int_s^t \mu(x)dx}$ with condition $p_{AA}(s,s) = 1$

$\Pr\left[R_S > w \mid X_S = i\right] = e^{-\int_s^{s+w} \lambda_i(t)dt}$

$$\Pr\left[X_s^{(+)} = j \mid X_s = i, R_s = w\right] = \frac{\mu_{ij}(s+w)}{\lambda_i(s+w)}$$

Integrated Form of Backward Qquation:

$$p_{ij}(s,t) = \Pr\left[X_t = j \mid X_s = i\right]$$
$$= \sum_{\ell \neq i} \int_0^{t-s} p_{\ell j}(s+w,t)\mu_{i\ell}(s+w)e^{-\int_s^{s+w}\lambda_i(u)du}dw$$

Probability of survival in Two-State Model:

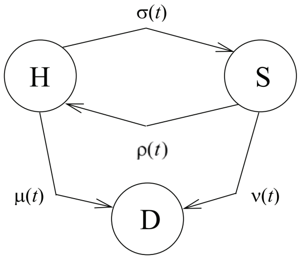
$$_w p_s = p_{AA}(s, s+w) = e^{-\int_s^{s+w}\mu(x)dx} = e^{-\int_0^w \mu(s+y)dy}$$

Integrated Form:

$$p_{ij}(s,t) = \sum_{\ell \neq i} \int_0^{t-s} p_{\ell j}(s+w,t)\mu_{i\ell}(s+w)e^{-\int_s^{s+w}\lambda_i(u)du}dw$$
(backward equation)

$$p_{ij}(s,t) = \sum_{k \neq j} \int_0^{t-s} p_{ik}(s,t-w)\mu_{kj}(t-w)e^{-\int_{t-w}^{t}\lambda_j(u)du}dw \text{ (forward equation)}$$

Transition Graph for Sickness and Death Model:



Probabilities:

$$\Pr\left[R_s > t-s \mid X_s = H\right] = e^{-\int_s^t(\sigma(u)+\mu(u))du}$$
$$\Pr\left[R_s > t-s \mid X_s = S\right] = e^{-\int_s^t(\rho(u)+v(u))du}$$

SURVIVAL MODELS

Complete Future Lifetime: $\mathbf{T}_x$ → The pdf of $\mathbf{T}_x$ is $f_x(t) = {}_tp_x\ \mu_{x+t}$.

Curtate Future Lifetime: $\mathbf{K}_x = \lfloor \mathbf{T}_x \rfloor$ → The pmf of $\mathbf{K}_x$ is $\Pr(\mathbf{K}_x = k) = {}_kp_x\ q_{x+k}$.

Cumulative Distribution Function: $F_x(t) = P\left[\mathbf{T}_x \leq t\right] = {}_tq_x$

Survival Function: $S_x(t) = P\left[\mathbf{T}_x > t\right] = 1 - F_x(t) = {}_tp_x = \dfrac{S(x+t)}{S(x)} = \dfrac{{}_{x+t}p_0}{{}_xp_0}$

Central Rate of Mortality: $m_x = \dfrac{q_x}{\int_0^1 {}_tp_x dt}$

Complete Expectation of Life: $E\left[\mathbf{T}_x\right] = \overset{o}{e}_x = \int_0^{\omega-x} {}_tp_x dt \cong e_x + 1/2$

Curtate Expectation of Life: $E\left[\mathbf{K}_x\right] = e_x = \sum_{k=1}^{[\omega-x]} {}_kp_x$

Complete Future Lifetime Variance: $\mathrm{Var}\left[\mathbf{T}_x\right] = \int_0^{\omega-x} t^2\,{}_tp_x\mu_{x+t} - \overset{o}{e}{}_x^2$

Curtate Future Lifetime Variance: $\mathrm{Var}\left[\mathbf{K}_x\right] = \sum_{k=0}^{[\omega-x]} k^2{}_kp_x q_{x+k} - e_x^2$

Force of Mortality: $\mu_x = \lim_{h\to 0^+} \dfrac{1}{h} \times P[\mathbf{T} \leq x+h \mid \mathbf{T} > x]$

Force of Mortality Condition: $\lim_{t\to\infty} \int_0^t \mu_s ds = \infty$

Force of Mortality Formulas:

$${}_tp_x = e^{-\int_0^t \mu_{x+s}ds} = e^{-\int_x^{x+t}\mu_s ds}$$
$${}_tq_x = \int_0^t {}_sp_x\ \mu_{x+s}ds$$
$${}_{t|u}q_x = \int_t^{t+u} {}_sp_x\ \mu_{x+s}ds$$

Exponential Model: $\mu_x = \mu$ ${}_tp_x = \exp(-\mu t)$

Weibull Model: ${}_tp_x = \exp\left[-\alpha t^\beta\right]$ $\mu_{x+t} = \alpha\beta t^{\beta-1}$

Gompertz' Law: $\mu_x = Bc^x$ ${}_tp_x = g^{c^x(c^t-1)}$ where $g = \exp\left(\frac{-B}{\log c}\right)$

Makeham's Law: $\mu_x = A + Bc^x$ ${}_tp_x = s^t g^{c^x(c^t-1)}$ where $g = \exp\left(\frac{-B}{\log c}\right)$ and $s = \exp(-A)$

ESTIMATING THE LIFETIME DISTRIBUTION

Right Censoring: Observations are cut short, e.g., a mortality investigation before deaths of some participants, so all we know is $T_i > C_i$, where C_i is the censoring time and T_i is the lifetime of individual i. [Most common censoring in actuarial applications.]

Left Censoring: The censoring mechanism prevents us from knowing when entry into a state occurred, e.g., discovery of a medical condition at a regular examination tells us only that onset occurred sometime prior to the examination.

Interval Censoring: We can only say the event of interest fell within an interval of time, e.g., we know a death occurred in a particular calendar year but do not know precisely when.

Random Censoring: Censor times are random

Type I Censoring: Censoring times are known in advance

Type II Censoring: Observation is continued until a pre-determined number of deaths has occurred

Non-Informative Censoring: The censoring times C_i give no information about the lifetimes T_i; in random censoring, a sufficient condition for non-informative censoring is that T_i, C_i be independent

Discrete Hazard Function: $\lambda_j = P\left[T = t_j \mid T \geq t_j\right]$ $(1 \leq j \leq k)$ $\hat{\lambda}_j = \frac{d_j}{n_j}$

Kaplan-Meier Estimator: $\hat{S}(t) = \prod_{t_j \leq t}\left(1 - \hat{\lambda}_j\right)$

Greenwood's Formula for KM:

$$\text{Var}[\tilde{F}(t)] \approx (1-\hat{F}(t))^2 \sum_{t_j \leq t} \frac{d_j}{n_j\left(n_j - d_j\right)}$$

$$\text{Var}(\tilde{S}(t)) \approx (\hat{S}(t))^2 \sum_{t_j \leq t} \frac{d_j}{n_j\left(n_j - d_j\right)}$$

Nelson-Aalen Estimator: $\hat{\Lambda}_t = \sum_{t_j \leq t} \frac{d_j}{n_j}$ $\hat{F}_t = 1 - \exp\left(-\hat{\Lambda}_t\right)$

Greenwood's Formula for NA: $\text{Var}\left[\tilde{\Lambda}_t\right] \approx \sum_{t_j \leq t} \frac{d_j\left(n_j - d_j\right)}{n_j^3}$

Likelihood Function: $L = \prod_{i=1}^{n} \mu^{\delta_i} \exp\left(-\mu t_i\right)$ with death and censored values

Survival Function estimate by chaining together estimates of Force of Mortality:

$$\hat{S}_x(m) = {}_m\hat{p}_x = \exp\left(-\sum_{j=0}^{m-1} \hat{\mu}_{x+j}\right)$$

PROPORTIONAL HAZARDS MODELS

PH model, Hazard Function: $\lambda\left(t, z_i\right) = \lambda_0(t) \cdot g\left(z_i\right)$ $\lambda\left(t, z_i\right) = \lambda_0(t) \cdot g\left(z_i, t\right)$

PH model, Gompertz Hazard: $\lambda_i\left(t, z_i\right) = c^t \exp\left(\beta z_i^T\right)$

Cox PH Model: $\lambda\left(t; z_i\right) = \lambda_0(t) \exp\left(\beta z_i^T\right)$ where $\lambda_0(t)$ is the baseline hazard

Partial Likelihood for Cox Model: $L(\beta) = \prod_{j=1}^{k} \frac{\exp\left(\beta z_j^T\right)}{\sum_{i \in R(t_j)} \exp\left(\beta z_i^T\right)}$

Efficient Score Function: $u(\beta)=\left(\frac{\partial \log L(\beta)}{\partial \beta_1},\ldots,\frac{\partial \log L(\beta)}{\partial \beta_p}\right)$

Likelihood Ratio Statistic: A model with p covariates, and another model with $p+q$ covariates which include the p covariates of the first model L_p and L_{p+q} are the maximised log-likelihoods of the first and second models respectively $\rightarrow$ statistic: $-2\left(L_p-L_{p+q}\right)\sim\chi^2(q)$

ESTIMATION IN THE MARKOV MODEL

Two-state Model Probabilities:

$$\begin{aligned}{}_{t+dt}p_x &= {}_tp_x \times P[\text{Alive at } x+t+dt \mid \text{Alive at } x+t] \\ &\quad + {}_tq_x \times P[\text{Alive at } x+t+dt \mid \text{Dead at } x+t] \\ &= {}_tp_x \times (1-\mu_{x+t}dt+o(dt))\end{aligned}$$

Define a random variable $\mathbf{D}_i=\begin{cases}1 \text{ if the } i^{th} \text{ life is observed to die}\\ 0 \text{ if the } i^{th} \text{ life is not observed to die}\end{cases}$ and Waiting time $\mathbf{V}_t=\mathbf{T}_i-a_i$

Joint Distribution of $(\mathbf{D}_i,\mathbf{V}_i)$: $f_i\left(d_i,v_i\right)=\begin{cases}{}_{b_i-a_i}p_{x+a_i} & (d_i=0)\\ {}_{v_i}p_{x+a_i}\mu_{x+a_i+v_i} & (d_i=1)\end{cases}=\exp\left(-\int_0^{v_i}\mu_{x+a_i+t}dt\right)\mu_{x+a_i+v_i}^{d_i}$

JPF of all the $(\mathbf{D}_i,\mathbf{V}_i)$: $\prod_{i=1}^{N} e^{-\mu v_i}\mu^{d_i}=e^{-\mu(v_1+\ldots+v_N)}\mu^{d_i+\ldots+d_N}=e^{-\mu v}\mu^d$

Maximum Likelihood Estimate for μ: $\hat{\mu}=d/v$ **Estimator:** $\tilde{\mu}=\mathbf{D}/\mathbf{V}$

Exact Result: $E\left[\mathbf{D}_i-\mu\mathbf{V}_i\right]=0$ $\quad$ $\operatorname{Var}\left[\mathbf{D}_i-\mu\mathbf{V}_i\right]=E\left[\mathbf{D}_i\right]$

Poisson Model Alive-Dead: $\mathrm{P}[\mathbf{D}=d]=\frac{e^{-\mu E_x^c}\left(\mu E_x^c\right)^d}{d!}$

Estimator: $\tilde{\mu}=\frac{\mathbf{D}}{E_x^c}$ $\quad$ $\mathrm{E}[\tilde{\mu}]=\mu$ $\quad$ $\operatorname{Var}[\tilde{\mu}]=\frac{\mu}{E_x^c}$

Estimators of Healthy-Sick-Dead: $\tilde{\mu}=\frac{\mathbf{D}}{\mathbf{V}}$, $\quad$ $\tilde{v}=\frac{\mathbf{U}}{\mathbf{W}}$, $\quad$ $\tilde{\sigma}=\frac{\mathbf{S}}{\mathbf{V}}$, $\quad$ $\tilde{\rho}=\frac{\mathbf{R}}{\mathbf{W}}$

EXPOSED TO RISK

Principle of Correspondence A life alive at time t should be included in the exposure at age x at time t if and only if, were that life to die immediately, he or she would be counted in the death data d_x at age x

Exact Integral of E_x^c: $E_x^c=\int_K^{K+N+1}P_{x,t}dt$

Trapezium Approximation of E_x^c: $E_x^c\cong\sum_{t=K}^{K+N}{}^{1}\!/_{2}\left(P_{x,t}+P_{x,t+1}\right)$

Different Definitions of Age:

Definition of x	*Rate interval*	*$\hat{\mu}$ estimates*	*$\hat{q}$ estimates*
age last birthday	$[x,x+1]$	$\mu_{x+1/2}$	q_x
age nearest birthday	$[x-1/2,x+1/2]$	μ_x	$q_{x-1/2}$
age next birthday	$[x-1,x]$	$\mu_{x-1/2}$	q_{x-1}

GRADUATION AND STATISTICAL TESTS

Approximate Asymptotic Distribution, D_x: $D_x\sim\text{Normal}\left(E_x^c\mu_x,E_x^c\mu_x\right)$

Approximate Asymptotic Distribution, $\tilde{\mu}_x$: $\tilde{\mu}_x\sim\text{Normal}\left(\mu_x,\frac{\mu_x}{E_x^c}\right)$

"Third Differences" Criterion for Smoothness of Graduation:

Be small in magnitude compared with the quantities themselves; and progress regularly

Deviation at age x: Actual deaths $-$ Expected deaths $=D_x-E_x^c\mu_x^s$ or $D_x-E_x^c\overset{o}{\mu}_x$

Standardised Deviation: $z_x = \dfrac{D_x - E_x^c \mu_x^s}{\sqrt{E_x^c \mu_x^s}}$ or $\dfrac{D_x - E_x^c \overset{o}{\mu}_x}{\sqrt{E_x^c \overset{o}{\mu}_x}}$

Standardised Deviations Test: $X = \sum_{\text{all intervals}} \dfrac{(\text{Actual} - \text{Expected})^2}{\text{Expected}}$

Signs Test: Test statistic: P = Number of z_x that are positive

Find the smallest value k for which: $\sum_{j=0}^{k} \binom{m}{j} \dfrac{1}{2}^m \geq 0.025$

$P \sim \text{Normal}(1/2m, 1/4m)$

Cumulative Deviations Test: $D_x \sim \text{Normal}\left(E_x^c \overset{o}{\mu}_x, E_x^c \overset{o}{\mu}_x\right)$ $\qquad \dfrac{\sum_{\text{all ages}} \left(D_x - E_x^c \overset{o}{\mu}_x\right)}{\sum_{\text{all ages}} E_x^c \overset{o}{\mu}_x} \sim \text{Normal}(0,1)$

Grouping of Signs Test: Test statistic: G = Number of groups of positive z_x's

Find the smallest value k for which: $\sum_{t=1}^{k} \dfrac{\binom{n_1 - 1}{t-1}\binom{n_2+1}{t}}{\binom{m}{n_1}} \geq 0.05$

$G \sim \text{Normal}\left(\dfrac{n_1(n_2+1)}{n_1+n_2}, \dfrac{(n_1 n_2)^2}{(n_1+n_2)^3}\right)$

Serial Correlations Test: Test statistic: $r_j = \dfrac{\sum_{i=1}^{m-j} \left(z_i - \bar{z}^{(1)}\right)\left(z_{i+j} - \bar{z}^{(2)}\right)}{\sqrt{\sum_{i=1}^{m-j} \left(z_i - \bar{z}^{(1)}\right)^2 \sum_{i=1}^{m-j} \left(z_{i+j} - \bar{z}^{(2)}\right)^2}}$

where $\bar{z}^{(1)} = \dfrac{1}{m-j} \sum_{i=1}^{m-j} z_i$ and $\bar{z}^{(2)} = \dfrac{1}{m-j} \sum_{i=1}^{m-j} z_{i+j}$

or $r_j \cong \dfrac{\sum_{i=1}^{m-j} (z_i - \bar{z})(z_{i+j} - \bar{z})}{\dfrac{m-j}{m} \sum_{i=1}^{m} (z_i - \bar{z})^2}$ where $\bar{z} = \dfrac{1}{m} \sum_{i=1}^{m} z_i$

$r_j \sqrt{m}$ can be tested against the Normal(0, 1) distribution

Too high a value indicates a tendency for deviations of the same sign to cluster

METHODS OF GRADUATION

Formulae: $\mu_x = (\text{polynomial}(1)) + \exp(\text{polynomial}(2))$

Graduation by Reference to Standard Table: $\overset{o}{\mu}_x = f\left(\mu_x^s\right)$

Natural Cubic Spline: $\mu_x = \alpha_0 + \alpha_1 x + \sum_{j=1}^{n} \beta_j \phi_j(x)$ where $n \geq 3$ and $\phi_j(x) = \begin{cases} 0 & x < x_j \\ (x - x_j)^3 & x \geq x_j \end{cases}$

Statistical Test Substitute the graduated estimates for the standard table quantities above and use the deviations $D_x - E_x^c \overset{o}{\mu}_x$

χ^2-statistic is unchanged, but we must reduce the number of degrees of freedom.

the cumulative deviations test cannot be used if the cumulative deviation is automatically zero because of the graduation procedure

MORTALITY PROJECTION

Reduction Factor: $R_{x,t} = \alpha_x + (1-\alpha_x)(1-f_{n,x})^{t/n}$

Mortality Rate: $q_{x,t} = R_{x,t} q_x$

Lee-Carter Model: $\ln m_{x,t} = a_x + b_x k_t + \epsilon_{x,t}$ and usual constraints are that $\sum_x b_x = 1$ and $\sum_t k_t = 0$

Actual Future Value: $\hat{k}_{t_0+l} = \hat{k}_{t_0} + l\hat{\mu} + \sum_{j=1}^{l} \hat{\epsilon}_{t_0+j}$

Predicted Future Mortality Rate:

$$\ln m_{x,t_0+l} = \hat{a}_x + \hat{b}_x \hat{k}_{t_0+l}$$

Age-period-cohort Extension:

$$\ln m_{x,t} = a_x + b_x^1 k_t + b_x^2 h_{t-x} + \epsilon_{x,t}$$

Expcted number of deaths under Spline Mortality Projection:

$$\ln\left[E\left(D_x\right)\right] = \ln E_x^c + \sum_{j=1}^{S} \theta_j B_j(x)$$

Method of P-splines: Specify the knot spacing and degree of the polynomials in each spline

Define a roughness penalty, $P(\theta)$, which increases with the variability of adjacent coefficients This, in effect, measures the amount of roughness in the fitted model.

Define a smoothing parameter, λ, such that if $\lambda = 0$, there is no penalty for increased roughness, but as λ increases, roughness is penalised more and more.

Estimate the parameters of the model, including the number of splines, by maximising the penalised log likelihood: $l_p(\theta) = l(\theta) - \lambda P(\theta)$

MACHINE LEARNING

Supervised Learning: Has a response variable that is influenced by explanatory variables

Unsupervised Learning: Relates the observations to each other or finds patterns in the observations

Prediction: Determining what the response will be for new sets of predictor values

Inference: Understanding how the explanatory variables influence the response

Mean Square Error (MSE): $MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2$

Bias: Measures the extent to which the expected value of the estimator differs from the true value

Variance: Measures how much the estimator varies with different random samples of data

Bias-Variance Trade-off: $E\left[\left(y_0 - \hat{f}(\boldsymbol{x}_0)\right)^2\right] = \mathrm{Var}(\epsilon_0) + \mathrm{Var}\left[\hat{f}(\boldsymbol{x}_0)\right] + \left\{f(\boldsymbol{x}_0) - E\left[\hat{f}(\boldsymbol{x}_0)\right]\right\}^2$

Confusion Matrix:

Patient actually has condition	Test result classified/predicts patient as having condition: Yes	No
Yes	True positive (TP)	False negative (FN)
No	False positive (FP)	True negative (TN)

Accuracy of Classifier: $\text{Accuracy} = \frac{\text{TN}+\text{TP}}{n}$

Precision of Classifier: $\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$

Recall of Classifier: $\text{Recall} = \dfrac{\text{TP}}{\text{TP} + \text{FN}} = \text{Sensitivity}$

F1 Score of Classifier: $\text{F1 score} = \dfrac{2 \times \text{ precision x recall}}{\text{precision } + \text{ recall}}$

Specificity of Classifier: $\text{Specificity} = \dfrac{\text{FP}}{\text{TN} + \text{FP}} = 1 - \text{False positive rate}$

Training Set: The subset of the available observations used for fitting or "training" the given statistical learning method

Validation Set: The fitted statistical learning model is tested out by making predictions for observations in the validation set

Cross-Validation: Select test data sets in a systematic non-random fashion, and then average the MSE results from all runs

K-fold Cross-Validation: $\text{CV}_{(k)} = \frac{1}{k} \sum_{i=1}^{k} \text{MSE}_i$

The data is randomly partitioned into k subsets, each subset as test data, the rest as training data

Feature Scaling: Center input variables by subtracting the sample mean and rescale each variable to have a standard deviation of 1

Akaike Information Criterion: $\text{AIC} = \text{deviance } + 2d$

Bayesian Information Criterion:

$\text{BIC} = \text{deviance} + \ln(n) d$

Ridge Regression Penalty: $g(\beta_0, \ldots, \beta_d) = \sum_{j=1}^{d} \beta_j^2$

Lasso Regression Penalty: $g(\beta_0, \ldots, \beta_d) = \sum_{j=1}^{d} |\beta_j|$

Gini Index: $\sum_{j=1}^{J} n_j \sum_{k=1}^{K} p_{jk}(1 - p_{jk})$

Regression Tree Error: $\sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_i - \widehat{y}_J)^2$

K-Means Clustering: group the cases in a data set into K disjoint clusters in which the cases are relatively homogeneous

K-Means Clustering Algorithm:

1. Split the cases arbitrarily into k clusters.
2. determine the ***centroid*** (using the mean values of the data points assigned to that cluster) of the kth cluster
3. Assign each case to the closest cluster in terms of **Euclidean distance**

$$\left(\text{dist}(\boldsymbol{x}, \boldsymbol{k}) = \sqrt{\sum_{j=1}^{J} (x_j - k_j)^2} \right)$$

4. Repeat steps 2–3 until cluster assignments do not change