

Probability

→ Distribution shortcuts

→ Exp dist → memoryless property
 $P(X > r+s) = P(X > r)P(X > s)$
 OR $P(X > r+s | X > s) = P(X > r)$

→ $E(X | X > c) = c + \theta$
 \downarrow
 $= \frac{\int_c^\infty x f(x) dx}{P(X > c)}$

→ If $Y \sim \text{Exp}(\lambda) \Rightarrow P(X_1 \oplus X_2) = \frac{2\theta}{2 + \theta}$
 $\Rightarrow Y \sim \text{min}(X_1, \dots, X_n) \sim \text{Exp}(\lambda = E(X))$

→ If $X \sim \text{Uniform}(a, b) \Rightarrow X | X > c \sim \text{Uniform}(c, b)$

→ If $X \sim \text{Pareto}(\alpha, \theta) \Rightarrow E(X | X > c) \sim \text{Pareto}(\alpha, \theta + c)$

→ Random starts → coefficient of variation $CV = \frac{\sigma}{\mu}$

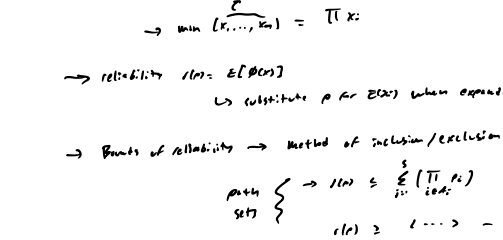
→ Poisson Processes

→ Compound Poisson Processes

→ $N \sim \text{Poisson}$
 $X \sim \text{Exp}(\lambda)$
 \hookrightarrow discrete or continuous
 $S(t) = \sum_{i=1}^N X_i$
 $\Rightarrow E(S(t)) = E(N)E(X)$
 $V(S(t)) = E(N)E(X^2)$
 \Rightarrow use normal approximation + continuity correction

→ Thinning → $N \sim \text{Poisson}$ → P_1, \dots, P_k are probabilities of some breakdown
 $\Rightarrow N_j(t) \sim \text{Poisson}(P_j t)$

→ $P(PA \text{ before } PB) = \text{Binomial, drop } n \text{ & } k \text{ down}$
 $P_i = \frac{\lambda_i}{\lambda_i + \lambda_j}$



→ Reliability Theory

→ Building structure functions

→ $\text{Max}(x_1, \dots, x_n) = 1 - \prod (1 - x_i)$
 \hookrightarrow binary

→ $\text{Min}(x_1, \dots, x_n) = \prod x_i$

→ reliability $r(t) = E(R(t))$

\hookrightarrow substitute p for $E(X)$ when expanded enough

→ Bounds of reliability → Method of inclusion/exclusion

$\left. \begin{matrix} \text{paths} \\ \text{sets} \end{matrix} \right\} \Rightarrow r(t) \leq \sum_{i=1}^s \prod_{j \in A_i} p_j$ \hookrightarrow multiply all in set, then add +
 $r(t) \geq 1 - \dots = 1 - \sum_{j \in B_i} \prod_{i \in A_i} (1 - p_i)$ \hookrightarrow multiply all in either set, then add +

OR
 $\left. \begin{matrix} \text{cut} \\ \text{sets} \end{matrix} \right\} \Rightarrow 1 - r(t) \leq \sum_{i=1}^s \prod_{j \in B_i} (1 - p_j)$
 \downarrow
 complement everywhere $1 - r(t) \geq 1 - \dots = 1 - \sum_{j \in B_i} \prod_{i \in A_i} (1 - p_i)$

→ Lifetime → $E(T) = \int_0^\infty r(t) dt$
 \hookrightarrow reliability w/ survival function reflecting λ

→ random signals → $P_n \approx 1 - n e^{-n}$

→ Gambler's ruin

→ $P(\text{win} | \text{bias}) = \begin{cases} 1 - \left(\frac{q}{p}\right)^i & \text{if } p \neq q \\ 1 - \left(\frac{q}{p}\right)^i & \text{if } p = q \end{cases}$
 \hookrightarrow if $p \neq q$
 \hookrightarrow if $p = q$

→ Life contingencies

→ $t P_x = \frac{P_{\text{surv}}}{P_x} = P_x P_{x+1} \dots P_{x+t}$
 $t q_x = 1 - t P_x$

→ Equivalence principle → $E(\text{Premium}) = E(\text{Benefit})$

\star discount correctly

→ premium → $t E_x = v^t \cdot P_x$

→ annuity due → $\ddot{a}_x = 1 + v P_x + v^2 P_x + \dots$

paying @ beginning of year for life \downarrow $\ddot{a}_x = 1 + v P_x \ddot{a}_{x+1}$

→ [I for annuities → "life bihumal"]

$\rightarrow n P + 2 n P \sqrt{n P q}$
 \hookrightarrow $e P q \rightarrow$ same for both first

→ Hazard functions

→ $h(x) = \frac{f(x)}{S(x)}$

→ Cumulative Hazard function → $H(x) = \int_0^x h(t) dt$
 (same relationship as pdf & cdf)

→ $S(x) = e^{-H(x)}$

→ Insurance applications

→ General Coverage modification

$E(Y^*) = H(1+r) [E(X + \frac{d}{1+r}) - E(X_{\text{and}})]$
 max covered loss: $\frac{\frac{d}{1+r} + d}{1+r}$

→ Continuity is after the fact

→ payment per loss & payment per payment

→ $E(Y^*) = \frac{E(Y)}{S(d)} \rightarrow E[(Y-d)^+] = E(Y) - E(X_{\text{and}})$
 $\hookrightarrow E[(Y-d)^+] = E(Y) \rightarrow$ when excess loss variable extra given is only difference to Y^*

→ $E[X_{\text{and}}] = \int_0^d x f(x) dx + u S(d)$

→ loss elimination ratio $LER = \frac{E(X_{\text{and}})}{E(Y)} = \% \text{ saved w/ deductible}$

→ Double expectation theorem

→ $E(X) = E_Y[E_X(Y|X)]$, $V(X) = E_Y[V_X(Y|X)] + V_Y[E_X(Y|X)]$

\Rightarrow use for mixtures (i.e. bihumal dist)

→ Tail properties of distributions

→ $CTE_q = E(X | X > \pi_q)$
 \downarrow
 $= \frac{\int_{\pi_q}^\infty x f(x) dx}{1-q}$
 $= \pi_q + \frac{E(X - \pi_q | X > \pi_q)}{1-q} = \frac{E(X) - E(X \wedge \pi_q)}{1-q}$
 $= \text{starting} + \text{contin}$

→ Tail weight → implies heavier tails

- 1) loss from numerator
- 2) $S_1(x) \rightarrow \pi$
- 3) $S_2(x) \rightarrow \pi$
- 4) upper CTE or π_q

→ Greedy algorithms → $E(\text{total cost}) = 0 \sum \frac{1}{i}$, where $C_{ij} = \text{Exp}(\theta)$

→ Simulation

→ Accept-reject if $\frac{f(x)}{g(x)} \geq u$
 \hookrightarrow acceptable enough
 \Rightarrow at least u

→ Discrete mixtures → $f_Y(y) = \sum_{i=1}^n w_i f_i(y) \rightarrow$ "mix pots"

same for pdfs & moments & (derivatives)

→ Gamma hierarchical models

→ Poisson-Gamma mixture → $\text{NB} (r=\alpha, \beta=\theta)$

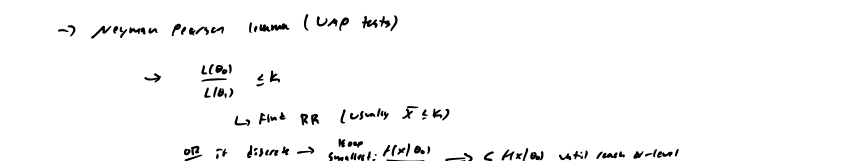
→ Exponential-inverse gamma → $\text{Pareto}(\alpha, \theta)$

→ Splines → $f_Y(y) = \begin{cases} C_1 h_1(y) \\ \vdots \\ C_n h_n(y) \end{cases} \quad 1 = \int_{-\infty}^\infty \text{pieces to find } C_i$

→ Markov chains

→ unconditional probs → $P(X_n = j) = \sum_{i=1}^n a_i P_{ij}^n$
 \hookrightarrow same for all n

→ Time reversibility → $P_{i,j} = \frac{\pi_i P_{ij}}{\pi_j}$



Statistics

→ Neyman Pearson (NAP tests)

→ $\frac{L(\theta_0)}{L(\theta_1)} \leq k$
 \hookrightarrow find RR (usually $F \leq k$)
 OR if discrete → $\frac{L(\theta_0)}{L(\theta_1)} \Rightarrow \leq \frac{L(\theta_0)}{L(\theta_1)}$ until reach α -level

→ Information

→ $J = -E(U') = E(U'')$
 \hookrightarrow same for $U_i \Rightarrow J = n E(U'')$

→ CR-LB = J^{-1}

→ Percentage matching

→ $j = (n+1)P$
 $\rightarrow X_{(j)} \Rightarrow e.g. \text{ and } yes$
 $i = 1, 2, 3, 4$
 $j = 4.7 \Rightarrow 0.8 X_{(4)} + 0.2 X_{(5)}$

→ Tests

→ Kolmogorov-Smirnov test



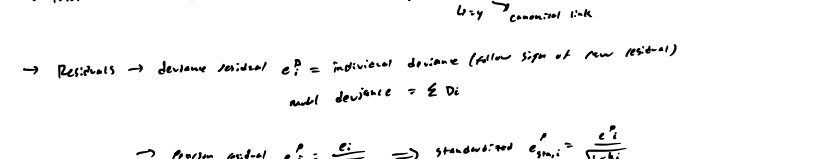
→ $\chi^2 - \text{GOF} \rightarrow \sum \frac{(O_i - E_i)^2}{E_i}$

→ Order statistics

→ $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$
 $F_Y(y) = \sum_{i=1}^n \binom{n}{i} (F_X(y))^i (1 - F_X(y))^{n-i}$
 \downarrow
 $\sim \text{Bin}(n, P(X \leq y)) + P(Y \leq y)$

→ Kernel density estimation

→ $\hat{f}(x) = \frac{1}{n} \sum K_h(x)$
 \hookrightarrow dim kernel centered at each x_i



GLMs

→ $R^2_{adj} = 1 - \left(\frac{n-1}{n-p} \right) \left(\frac{SSR}{SSR + SSR_0} \right)$

→ Principal components

→ $\sum \beta_{ij}^2 = 1$ & $\beta_{ij} \cdot \beta_{ik} = 0$
 dot product = 0
 $\rightarrow Z_i = \sum \beta_{ij} X_{ij}$
 \hookrightarrow PC score

→ Wald test → $\frac{\hat{\beta} - \beta}{\sqrt{V(\hat{\beta})}} \sim \chi^2_p$
 \hookrightarrow additional # of parameters
 OR regular z test

→ Local regression → minimize $\sum K_h(X_i - X_j)^2$
 $\hookrightarrow X_i \cdot Y_i$

→ F-test → * Relative to full model

→ Deviance = $2 [l_{\text{full}} - l_{\text{reduced}}]$
 \hookrightarrow where $\beta_i = 0$ (parameter not used)

→ Trivia → unknown glm = canonical link \Rightarrow exp { $\beta_1 \ln(\text{mean}) + \text{const} + \text{link}$ }
 \hookrightarrow canonical link

→ Residuals → deviance residual $e_i^D = \text{individual deviance (follow sign of new residual)}$
 model deviance = $\sum D_i^2$

→ Pearson residual $e_i^P = \frac{e_i}{\sqrt{V(\hat{\mu}_i)}} \Rightarrow$ standardized $e_{\text{std},i} = \frac{e_i^P}{\sqrt{1 - h_i}}$
 \rightarrow Pearson chi-sq stat = $\sum (e_i^P)^2$

→ VIF = $\frac{1}{1 - R_j^2}$ → $S =$ multicollinearity
 $\hookrightarrow X_j \sim X_{-j}$

→ Leverage $h_i = \frac{1}{n} + \frac{\sum x_{ij}^2}{\sum x_{ij}^2}$ → $\frac{1}{n} < h_i < 1$, $\text{diag}(H)$, $h_i = (n+1)$

→ Cook's distance $d_i = \frac{e_i^2 h_i}{MSE (n-1)(1-h_i)^2} = \frac{D_i^2 (n+1)}{(n-1)}$
 D_i^2 = function of d_i

→ LOOCV = $\frac{1}{n} \sum \left(\frac{e_i}{1 - h_i} \right)^2$

→ H of unique models → forward/backward selection = $(1 + \frac{1}{n}) \frac{1}{n}$
 \rightarrow best subset = 2^n

→ Weighted least squares → $\hat{\beta}^{OLS} = (X' W X)^{-1} X' W y$
 \hookrightarrow sandwich W_i & repeat y w/o

→ splines → first derivatives are equal (if quadratic spline)

→ Analysis $\sigma^2 = \frac{1}{n} (SSE + p \cdot MSE_0)$

→ Binary regression → complementary log-log link → $\ln(-\ln(1-p)) = x^T \beta$