Probability

→ Distribution shortcuts

→ Exp distrib memoryless property

$P(X > x+s) = P(X>x)P(X>s)$

⟺ $P(X > x+s \mid X>s) = P(X>x)$

→ $E(X \mid X>s) = s + \theta$ [illegible]

→ [illegible]

→ If $X \sim \text{uniform}(a,b)$ ⇒ [illegible]

→ If $X \sim \text{Poisson}$ [illegible]

→ [illegible] coefficient of variation $CV = \frac{\sigma}{\mu}$

→ Poisson Processes

→ Compound Poisson Process

$N \sim \text{Poisson}$, $Y \sim$ [illegible], $S(t) = \sum_{i=1}^{N} Y_i$

⇒ $E(S) = E(N)E(Y)$

$V(S) = E(N)E(Y^2)$

→ Thinning [illegible]

→ [illegible]

→ Reliability theory

→ Building structure functions

→ $\max(x_1, \ldots, x_n) = 1 - \prod(1-x_i)$

→ $\min(x_1, \ldots, x_n) = \prod x_i$

→ Bounds of reliability → method of inclusion/exclusion

→ $E(T) = \int_0^\infty S(t)\,dt$

→ random graphs → $P_n = 1 - q^{n-1}$

→ Gamblers ruin

$P(\text{win} \mid X(0)) = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^N} & \text{if } p \neq 0.5 \\ \frac{i}{N} & \text{if } p = 0.5 \end{cases}$

→ Life contingencies

→ ${}_tp_x = \frac{l_{x+t}}{l_x}$

${}_tq_x = 1 - {}_tp_x$

→ [illegible]

→ Hazard function

→ $h(x) = \frac{f(x)}{S(x)}$

→ Cumulative Hazard function → $H(x) = \int_{-\infty}^{x} h(t)\,dt$

(same relationship as pdf + cdf)

→ $S(x) = e^{-H(x)}$

→ Insurance applications

→ General coverage modification

$E(Y^L) = \alpha(1+r)\left[E\left(X \wedge \frac{u}{1+r}\right) - E(X \wedge d)\right]$

↳ $\frac{d}{1+r}$

Max covered loss: $\frac{\frac{u}{\alpha}+d}{1+r}$

→ Coinsurance is after the fact

→ payment per loss & payment per payment

→ $E(Y^P) = \frac{E(Y^L)}{S(d)}$ → $E[(X-d)_+] = E(X) - E(X \wedge d)$

↳ $= E(X-d \mid \underbrace{X>d}) = e(d)$ → mean excess loss variable

extra given is only difference to Y^L

→ $E(X \wedge u) = \int_0^u x f(x)\,dx + u\,S(u)$

→ Loss elimination ratio $LER = \frac{E(X \wedge d)}{E(X)}$ = % saved w/ deductible

→ Double expectation theorems

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

⇒ use for mixtures (ie hierarchical dists)

→ Tail properties of distributions

→ $CTE_q = E(X \mid X > \pi_q)$

$= \frac{\int_{\pi_q}^{\infty} x f(x)\,dx}{1-q}$

$= \pi_q + \underbrace{E(X-\pi_q \mid X>\pi_q)}$ ↳ $= \frac{E(X) - E(X \wedge \pi_q)}{1-q}$

$=$ <starting point> + <extra>

→ Tail weight → implies heavier tails

1) less raw moments

2) $\frac{S_1(x)}{S_2(x)} \to \infty$

3) $h(x) \downarrow$ as $x \uparrow$

4) larger CTE_q or π_q

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

→ Simulation

→ Accept-reject if $\frac{f(x)}{c\,g(x)} \geq u$

↳ acceptable enough

⇒ At least u

→ Discrete mixtures → $f_Y(y) = \sum_{i=1}^{n} w_i f_{X_i}(y)$ → "mix pdfs"

<same for cdfs & moments>

(& survivals)

→ Common Hierarchical models

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

→ Exponential-inverse gamma ~ $Pareto(\alpha, \theta)$

→ Splices → $f_Y(y) = \begin{cases} c_1 f_{X_1}(y) \\ \vdots \\ c_n f_{X_n}(y) \end{cases}$ $1 = \int_{-\infty}^{\infty}$ pieces to find c_i

→ Markov chains

→ unconditional probs → $P(X_n = j) = \sum_{i=1}^{\infty} \alpha_i P_{ij}^{n}$

↳ time 0 prob

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

Statistics

→ Neyman Pearson lemma (UMP tests)

→ $\frac{L(\theta_0)}{L(\theta_1)} \leq k$

↳ Find RR (usually $\bar{X} \leq k$)

OR if discrete → keep smallest: $\frac{f(x|\theta_0)}{f(x|\theta_1)}$ ⇒ $\sum f(x|\theta_0)$ until reach α-level

→ Information

→ $\underbrace{I = -E(U') = E(U^2)}$

↳ solve for U_i ⇒ $I = n(-E(U_i'))$

→ $CR\text{-}LB = I^{-1}$

→ Percentage matching

→ $j = (n+1)p$

→ $X_{(j)}$ ⇒ e.g.) drop 4 & 5

$i = 1, 2, 3, 6$

$j = 4.2$ ⇒ $0.8X_{(3)} + 0.2X_{(6)}$

→ Tests

→ Kolmogorov Smirnov test

x	fitted cdf @ x	empirical cdf	D_j
1	a	e	a
5	b	f	max(|b-e|, |b-f|)
7	c	g	⋮
9	d	h	

→ χ^2-GOF → $\sum \sum \frac{(o-e_i)^2}{e}$

→ Order Statistics

→ $X_i \overset{iid}{\sim} f(x)$ & $Y = X_{(j)}$

$F_Y(y) = \sum_{i=j}^{n} \binom{n}{i} (F_X(y))^i (1-F_X(y))^{n-i}$

↓ $\sim Bin(n, P(X \leq y))$ & $P(Y \geq j)$

→ Kernel density estimation

→ $\hat{f}(x) = \frac{1}{n} \sum K_i(x)$

↳ draw Kernel centered at each X_i

GLMs

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

→ Principal components

→ $\sum_{j=1}^{p} \phi_{1,j}^2 = 1$ & $\phi_1 \cdot \phi_2 = 0$

dot product $= \sum a_i b_i$

→ $z_i = \sum \phi_{h,j} x_{i,j}$

↳ PC score

→ Wald test → $\frac{\hat{\beta} - \beta}{V(\hat{\beta})} \sim \chi^2$

↳ additional # of parameters

OR

regular z test

→ local regression → minimize $\sum K_i (y_i - \hat{y}_i)^2$

↳ $x_i^T \hat{\beta}$

→ F-test → * Relative to full model

→ Deviance $= 2[\ell_{sat} - \ell_{fit}]$

↳ uses $\hat{y}_i = y_i$ (perfectly fit model)

→ Trivia → unbiased glm = canonical link ⇒ $\exp\{ \ln(\underbrace{a(y)}_{=y} \underbrace{b(\theta)} + c(\theta) + d(y) \}$ canonical link

→ Residuals → deviance residual e_i^D = individual deviance (follow sign of raw residual)

model deviance $= \sum D_i$

→ Pearson residual $e_i^P = \frac{e_i}{\sqrt{V(\hat{Y}_i)}}$ ⇒ standardized $e^P_{stand,i} = \frac{e_i^P}{\sqrt{1-h_i}}$

→ pearson chi-sq stat $= \sum (e_i^P)^2$

→ $VIF = \frac{1}{1-R_j^2}$ > 5 = problematic

↳ $X_j \sim X_{-j}$

→ Leverage $h_i = \frac{1}{n} + \frac{S_{xx_i}}{S_{xx}}$ → $\frac{1}{n} < h_i < 1$, diag(H), $\sum h_i = (p+1)$

→ Cooks distance $d_i = \frac{e_i^2 h_i}{MSE(p+1)(1-h_i)^2} = \frac{DFits^2}{(p+1)}$

DFits = function of d_i

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

→ # of unique models → forward/backward selection $= 1 + \frac{p(p+1)}{2}$

→ best subsets $= 2^p$

→ weighted least squares → $\hat{\beta}^{(i)} = (X'WX)^{-1\,(i-1)}(X'Wz)^{(i-1)}$

sandwich Ws & replace y w/ z

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

→ Mallows $C_p = \frac{1}{n}(SSE + 2p \cdot MSE_?)$

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