FreeFellow

FORMULA SHEET

Exam MAS-I

CAS · Modern Actuarial Statistics I

103 FORMULAS

3 TOPICS

freefellow.org/cas-masi/formulas

Gamma waiting time density for the n-th arrival
$f_{S_n}(s) = \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!}$ — λ = rate, n = event index, s = waiting time, mean n/λ, variance n/λ²

Compound Poisson mean and variance
$E[S(t)] = \lambda t\, E[X],\ \mathrm{Var}(S(t)) = \lambda t\, E[X^2]$ — λ = rate, t = time, X = iid severity, E[X²] = Var(X)+(E[X])²

NHPP mean function over an interval
$\Lambda(a,b) = \int_a^b \lambda(s)\,ds$ — λ(s) = time-varying intensity, [a,b] = interval; equals both mean and variance of N(b)-N(a)

Monte Carlo sample size for target half-width
$n \approx (1.96\,s/h)^2$ — s = pilot sample SD, h = target half-width, n = required number of draws

Exponential inversion from a uniform
$X = -\frac{1}{\lambda}\ln(1-U)$ — U = Uniform(0,1) draw, λ = exponential rate, X = exponential severity draw

Monte Carlo 95% confidence interval half-width
$\hat{\theta}_n \pm 1.96\,s/\sqrt{n}$ — \hat\theta_n = sample mean of g(X_i), s = sample SD, n = independent draws

Inversion method draw from a uniform
$X = F^{-1}(U)$ — U = Uniform(0,1) draw, F^{-1} = generalized inverse CDF, X = draw from target distribution F

Reversionary annuity to y after x dies
$\bar{a}_{y|x} = \bar{a}_y - \bar{a}_{xy}$ — $\bar{a}_y$ = single-life annuity on y, $\bar{a}_{xy}$ = joint-life annuity

Joint-life survival under common shock
${}_tp_{xy} = {}_tp_x^* \cdot {}_tp_y^* \cdot e^{-\lambda t}$ — ${}_tp^*$ = private survival, λ = shared hazard rate, t = time

Survival function from cumulative hazard
$S(t) = \exp\left(-\int_0^t h(s)\,ds\right) = e^{-H(t)}$ — H(t) = cumulative hazard, h(s) = hazard rate, S(t) = survival probability

Last-survivor survival probability (inclusion-exclusion)
${}_tp_{\overline{xy}} = {}_tp_x + {}_tp_y - {}_tp_{xy}$ — ${}_tp_x$ = prob x survives t, ${}_tp_{xy}$ = joint survival

Mean residual life at age t
$e(t) = \frac{\int_t^\infty S(u)\,du}{S(t)}$ — S = survival function, t = current age, e(t) = expected remaining lifetime given survival to t

Joint-life continuous annuity under constant force and interest
$\bar{a}_{xy} = \dfrac{1}{\mu_x + \mu_y + \delta}$ — μ_x, μ_y = constant forces of mortality, δ = force of interest

Conditional survival probability (t-p-s)
${}_tp_s = \frac{S(s+t)}{S(s)}$ — S = survival function, s = current age, t = additional years survived

Hazard rate from density and survival
$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\ln S(t)$ — f = density, S = survival function, h = instantaneous failure rate

Constant-force whole life insurance EPV
$\bar{A}_x = \frac{\mu}{\mu+\delta}$ — μ = constant force of mortality, δ = constant force of interest

Whole life insurance EPV (discrete)
$A_x = \sum_{k=0}^{\infty} v^{k+1}\,{}_{k|}q_x$ — v = 1/(1+i), ${}_{k|}q_x$ = prob of death in year k+1 for life age x

Insurance-annuity fundamental identity (discrete)
$A_x = 1 - d\,\ddot{a}_x$ — d = i/(1+i) effective discount rate, $\ddot{a}_x$ = whole life annuity-due EPV, A_x = whole life insurance EPV

Whole life annuity-due EPV
$\ddot{a}_x = \sum_{k=0}^{\infty} v^k\,{}_kp_x$ — v = 1/(1+i), ${}_kp_x$ = prob life age x survives k years

Waiting time to the n-th event in a Poisson process
$S_n = \sum_{i=1}^n T_i \sim \mathrm{Gamma}(n,\lambda),\ E[S_n] = n/\lambda$ — T_i = iid Exponential(λ) gaps, n = event number, λ = rate

Non-homogeneous Poisson process mean function
$N(b) - N(a) \sim \mathrm{Poisson}\left(\int_a^b \lambda(s)\,ds\right)$ — λ(s) = intensity function, [a,b] = time interval

Poisson process count probability
$P(N(t)=k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ — λ = rate, t = interval length, k = number of events

Compound Poisson aggregate loss mean and variance
$E[S(t)] = \lambda t\, E[X]$, $\mathrm{Var}(S(t)) = \lambda t\, E[X^2]$ — λ = rate, t = time, X = claim severity

Fundamental matrix of an absorbing Markov chain
$N = (I-Q)^{-1}$ — I = identity, Q = transient-to-transient block of P, N_{ij} = expected visits to transient state j starting from i

Series system reliability with independent components
$R_s = \prod_{i=1}^n R_i$ — R_i = reliability of component i, n = number of components in series

Bridge reliability by conditioning on the center element
$R_{\text{bridge}} = R_C \cdot R_{\text{works}} + (1-R_C)\cdot R_{\text{fails}}$ — R_C = bridge element reliability, R_works = system reliability given C works, R_fails = given C fails

Parallel system reliability with independent components
$R_p = 1 - \prod_{i=1}^n (1-R_i)$ — R_i = reliability of component i, 1 - R_i = unreliability, n = number of parallel components

Limited expected value (survival form)
$E[X \wedge u] = \int_0^u S(x)\,dx$ — X = non-negative loss, u = cap/limit, S(x) = survival function 1-F(x)

Loss elimination ratio for ordinary deductible

$\mathrm{LER}(d) = E[X \wedge d]/E[X]$ — d = ordinary deductible, X = ground-up loss severity

Expected layer cost between deductible d and limit u

$E[\min(X, u) - \min(X, d)] = E[X \wedge u] - E[X \wedge d]$ — X = loss, d = attachment, u = exhaustion point

Exponential limited expected value

$E[X \wedge u] = \theta(1 - e^{-u/\theta})$ — θ = exponential mean, u = policy limit/cap

Collective risk model aggregate loss
$S = \sum_{i=1}^{N} X_i$ — N = random claim count, X_i = iid severities independent of N

Compound distribution variance
$\mathrm{Var}(S) = E[N]\,\mathrm{Var}(X) + \mathrm{Var}(N)\,E[X]^2$ — N = claim count, X = severity, S = aggregate loss

Panjer recursion for aggregate loss pmf
$f_S(sh) = \frac{1}{1-af_X(0)} \sum_{y=1}^{s} (a + b\,y/s)\,f_X(yh)\,f_S((s-y)h)$ — (a,b) = (a,b,0) class parameters, h = grid step

Normal approximation stop-loss premium
$E[(S-d)_+] = \sigma[\phi(z) - z(1-\Phi(z))]$ — z = (d - E[S])/σ, σ = SD of S, φ = standard normal pdf, Φ = cdf

MLE of the exponential rate parameter
$\hat{\lambda} = n/\sum X_i = 1/\bar{X}$ — n = sample size, ΣX_i = sufficient statistic, X̄ = sample mean

Exponential family canonical density
$f(x;\theta) = h(x)\,c(\theta)\exp\left(\sum_{j=1}^{k} w_j(\theta)\,t_j(x)\right)$ — h, c base functions; w natural params; t sufficient kernels

Fisher-Neyman factorization theorem
$f(x_1,\ldots,x_n;\theta) = g(T(x),\theta)\cdot h(x)$ — T = sufficient statistic, g depends on θ through T, h depends only on data

UMVUE of theta for Uniform(0, theta)
$\hat{\theta}_{\mathrm{UMVUE}} = \frac{n+1}{n}\max X_i$ — n = sample size, max X_i = largest order statistic (complete sufficient stat)

One-sample z test statistic for a mean
$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ — X-bar = sample mean, μ_0 = hypothesized mean, σ = population SD, n = sample size

Two-sided p-value for a z test
$p = 2\cdot P(|Z| \geq |z_{\mathrm{obs}}|)$ — z_obs = observed test statistic, probability computed under H_0

Likelihood ratio test statistic
$-2\ln(L_0/L_1) \sim \chi^2_k$ — L_0 = restricted-model likelihood, L_1 = full-model likelihood, k = number of restrictions

Sample size for target power in a one-sided z test
$n = \frac{\sigma^2(z_{1-\alpha}+z_{1-\beta})^2}{(\mu_0-\mu_1)^2}$ — σ = SD, α = Type I rate, β = Type II rate, μ_0 = null mean, μ_1 = alternative mean

Likelihood contribution under left-truncation and right-censoring
$L_i(\theta) = \dfrac{f(y_i)^{\delta_i} S(y_i)^{1-\delta_i}}{S(d_i)}$ — f = density, S = survival, δ_i = censoring indicator, d_i = left-truncation point

Nelson-Aalen cumulative hazard estimator
$\hat{H}(t) = \sum_{t_j \leq t} s_j/n_j$ — s_j = events at time t_j, n_j = risk-set size (counts only those with $d_i < t_j \leq y_i$)

Conditional density under left-truncation at a deductible
$f_{X|X>d}(x) = f(x)/S(d)$ for $x > d$ — f = unconditional density, S(d) = survival at deductible d

Likelihood for a right-censored sample
$L(\theta) = \prod_{i=1}^{n} f(y_i)^{\delta_i} S(y_i)^{1-\delta_i}$ — f = density, S = survival, δ_i = 1 if uncensored, 0 if right-censored at y_i

Likelihood contribution with left truncation and right censoring
$L_i(\theta) = [f(x_i \mid \theta)/S(d_i \mid \theta)]^{\delta_i} [S(u_i \mid \theta)/S(d_i \mid \theta)]^{1-\delta_i}$ — δ=1 if observed, d=truncation, u=censoring point

Fisher information for a single observation
$I(\theta) = E\left[(\partial \log f/\partial\theta)^2\right] = -E\left[\partial^2 \log f/\partial\theta^2\right]$ — f = density, θ = parameter

Cramer-Rao lower bound for an unbiased estimator
$\mathrm{Var}(\hat{\theta}) \geq 1/[nI(\theta)]$ — n = sample size, I(θ) = Fisher information per observation, θ = parameter

Mean squared error decomposition
$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [\mathrm{Bias}(\hat{\theta})]^2$ — Bias(θ̂) = E[θ̂] - θ, Var = sampling variance of the estimator

CDF of the sample minimum
$F_{(1)}(x) = 1 - [1 - F(x)]^n$ — F = parent CDF, n = sample size, X_{(1)} = minimum

Unbiased sample variance with Bessel's correction
$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ — n = sample size, X_i = i-th observation, $\bar{X}$ = sample mean, S^2 unbiased for σ^2

Computational shortcut for sum of squared deviations
$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$ — n = sample size, X_i = i-th observation, $\bar{X}$ = sample mean

Density of the k-th order statistic
$f_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} F(x)^{k-1}[1-F(x)]^{n-k} f(x)$ — F = CDF, f = pdf, n = sample size, k = rank

CDF of the sample maximum
$F_{(n)}(x) = [F(x)]^n$ — F = parent CDF, n = sample size, X_{(n)} = maximum

Standard error of the sample mean
$\mathrm{SE}(\bar{X}) = S/\sqrt{n}$ — S = sample standard deviation, n = sample size; uses σ/√n when σ known

Sample mean
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ — n = sample size, X_i = i-th observation, $\bar{X}$ = sample mean (unbiased for μ)

Uniform order statistic as a Beta distribution
$U_{(k)} \sim \mathrm{Beta}(k,\, n-k+1)$, $E[U_{(k)}] = k/(n+1)$ — n = sample size, k = rank, U = Uniform(0,1) sample

Z-test statistic for a single mean with known variance

$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ — $\bar{X}$ = sample mean, μ_0 = hypothesized mean, σ = known population SD, n = sample size

T-test statistic for a single mean with unknown variance

$T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$ — $\bar{X}$ = sample mean, μ_0 = hypothesized mean, s = sample SD, n = sample size

F-test statistic for the ratio of two variances

$F = \frac{s_1^2}{s_2^2} \sim F_{n_1-1,\, n_2-1}$ — s_1^2, s_2^2 = sample variances (larger on top), n_1, n_2 = sample sizes

Chi-square test statistic for a single variance

$W = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$ — n = sample size, s^2 = sample variance, σ_0^2 = hypothesized variance

Lognormal mean and variance

$E[X] = e^{\mu+\sigma^2/2},\ \mathrm{Var}(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$ — μ, σ = mean and SD of log X

Aggregate loss mean and variance (general two-term)

$E[S] = E[N]\,E[X],\ \mathrm{Var}(S) = E[N]\,\mathrm{Var}(X) + \mathrm{Var}(N)\,E[X]^2$ — N = claim count, X = iid severity

Negative binomial mean and variance

$E[N] = r\beta,\ \mathrm{Var}(N) = r\beta(1+\beta)$ — r = shape, β = scale; variance exceeds mean by factor (1+β)

(a,b,0) class recursion

$p_k/p_{k-1} = a + b/k$ — p_k = probability of k claims, a and b = family-specific constants (Poisson, NegBin, Binomial)

Pearson chi-square dispersion estimate

$\hat{\phi} = X^2/(n-p)$ where $X^2 = \sum (r_i^P)^2$ — n = sample size, p = parameter count, r_i^P = Pearson residual

Deviance of a GLM

$D = 2[\ell(\mathbf{y};\mathbf{y}) - \ell(\hat{\boldsymbol{\mu}};\mathbf{y})]$ — l(y;y) = saturated log-likelihood, l(μ̂;y) = fitted log-likelihood

Pearson residual for a GLM

$r_i^P = (y_i - \hat{\mu}_i)/\sqrt{V(\hat{\mu}_i)}$ — y_i = observed, μ̂_i = fitted mean, V(μ̂_i) = variance function at μ̂_i

McFadden pseudo R-squared

$R^2_{\text{McF}} = 1 - \ell_{\text{model}}/\ell_{\text{null}}$ — l_model = fitted log-likelihood, l_null = intercept-only log-likelihood

Incremental pure-premium GLM with base-rate offset

$\ln E[P_i] = \ln(B_i) + \beta_0 + \sum_j \beta_j x_{ij}$ — P = pure premium, B = current base premium, β = log-relativities to the base

Annualized Poisson claim frequency from an exposure-offset model

$\hat{\lambda}_i = \mu_i/E_i = \exp(\beta_0 + \sum_j \beta_j x_{ij})$ — λ = per-exposure rate, μ = expected count, E = earned exposure

Population-averaged prediction across a control variable

$\bar{\mu} = \sum_k p_k\, g^{-1}(\eta_k)$ — p_k = population share of control level k, η_k = linear predictor at level k, g = link function

Linear predictor in a GLM with an offset term

$\eta_i = o_i + \beta_0 + \sum_j \beta_j x_{ij}$ — o = known offset (coef fixed at 1), β = estimated coefficients, x = predictors, η = linear predictor

Score equation under the canonical link

$X^\top(y-\mu) = 0$ — X = design matrix, y = response vector, μ = fitted mean vector at the MLE

GLM response variance with dispersion and exposure weight

$\text{Var}(Y_i) = \phi\, V(\mu_i)/w_i$ — φ = dispersion, V = variance function, μ = mean, w = exposure weight

Log-link GLM with exposure offset

$\ln \mu_i = \ln(\text{exposure}_i) + x_i^\top \beta$ — μ = mean response, exposure = policy-years at risk, x = covariates, β = coefficients

Exponential family density form

$f(y;\theta,\phi) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y,\phi)\}$ — θ = canonical parameter, φ = dispersion, b = cumulant function, a,c = known functions

Likelihood ratio statistic for nested GLMs

$\Lambda = 2(\ell_1 - \ell_0) \tilde{\ } \chi^2_{\Delta p}$ — l₁ = full-model log-likelihood, l₀ = reduced-model log-likelihood, Δp = extra parameters

Akaike information criterion for GLM selection

$\text{AIC} = -2\ell + 2p$ — l = maximized log-likelihood, p = number of fitted parameters; lower is better

Elastic net penalized GLM objective

$\hat{\beta} = \arg\min_\beta \{-\ell(\beta) + \lambda[\alpha\|\beta\|_1 + (1-\alpha)\|\beta\|_2^2]\}$ — λ = penalty strength, α = L1/L2 mix (1 = lasso, 0 = ridge)

Bayesian information criterion for GLM selection

$\text{BIC} = -2\ell + p\ln n$ — l = maximized log-likelihood, p = parameters, n = sample size; lower is better

Extended linear model linear predictor and link

$\eta = \beta_0 + \sum_{j=1}^p \beta_j x_j,\ g(\mu) = \eta$ — η = linear predictor, β = coefficients, x = design columns, g = link, μ = mean response

Continuous-by-continuous interaction model

$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3(x_1 x_2)$ — x1, x2 = continuous predictors, β3 = interaction coefficient measuring departure from additivity

Degrees of freedom for a categorical-by-categorical interaction

$df_{int} = (k_1 - 1)(k_2 - 1)$ — k1, k2 = number of levels in the two categorical predictors; added on top of (k1-1)+(k2-1) main-effect df

Effective slope on x1 under a continuous interaction

$\partial\eta/\partial x_1 = \beta_1 + \beta_3 x_2$ — β1 = main-effect slope, β3 = interaction coefficient, x2 = partner predictor value

Standardized residual

$r_i = \dfrac{e_i}{s\sqrt{1-h_{ii}}}$ — e_i = raw residual, s = residual std error, h_ii = leverage of point i

Added variable plot residuals for predictor X_j

$e_{y|-j} = y - \hat{y}_{(-j)},\ e_{j|-j} = X_j - \hat{X}_{j,(-j)}$ — hats with (-j) = fitted values from regressions that exclude X_j

Hat matrix for ordinary least squares

$H = X(X^TX)^{-1}X^T$ — X = design matrix; h_ii = i-th diagonal of H is the leverage of observation i

Cook's distance for an observation

$D_i = \dfrac{r_i^2}{p+1} \cdot \dfrac{h_{ii}}{1-h_{ii}}$ — r_i = standardized residual, h_ii = leverage, p = number of predictors

Deviance-based pseudo R-squared for a GLM

$R^2_{\text{dev}} = 1 - \frac{D_{\text{model}}}{D_{\text{null}}}$ — D_model = deviance of fitted GLM, D_null = deviance of intercept-only model

Adjusted R-squared for linear regression

$R^2_{\text{adj}} = 1 - (1-R^2)\,\frac{n-1}{n-k-1}$ — n = sample size, k = number of slope parameters (intercept excluded)

Coefficient of determination for OLS with intercept

$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = \frac{\text{SSR}}{\text{SST}}$ — SSE = error sum of squares, SST = total sum of squares, SSR = regression sum of squares

Scaled deviance of a GLM from log-likelihoods

$D^* = 2(\ell_{\text{sat}} - \ell_{\text{model}})$, with unscaled $D = \phi D^*$ — l_sat = saturated log-likelihood, l_model = fitted log-likelihood, φ = dispersion

Freedman-Diaconis rule for histogram bin width

$h = 2 \cdot IQR/n^{1/3}$ — h = bin width, IQR = interquartile range of the data, n = sample size

Interquartile range

$IQR = Q_3 - Q_1$ — Q1 = first quartile (25th percentile), Q3 = third quartile (75th percentile)

Tukey upper outlier fence for a box plot

$\text{Upper fence} = Q_3 + 1.5 \cdot IQR$ — Q3 = third quartile, IQR = interquartile range; points above are flagged outliers

Tukey lower outlier fence for a box plot

$\text{Lower fence} = Q_1 - 1.5 \cdot IQR$ — Q1 = first quartile, IQR = interquartile range; points below are flagged outliers

F statistic for analysis of deviance with estimated dispersion

$F = (\Delta D / \Delta df) / \hat{\phi}$ — ΔD = deviance reduction from added terms, Δdf = added parameters, φ̂ = estimated dispersion

Wald z statistic for a GLM coefficient

$z = \hat{\beta}_j / \text{SE}(\hat{\beta}_j)$ — β̂_j = MLE of coefficient j, SE = standard error of the estimate

GLM deviance

$D = 2[\ell(\text{saturated}) - \ell(\hat{\beta})]$ — l = log-likelihood, saturated = one parameter per observation, β̂ = fitted MLE

Pearson chi-square goodness-of-fit statistic for a GLM

$X^2 = \sum (y_i - \hat{\mu}_i)^2 / V(\hat{\mu}_i)$ — y_i = observation, μ̂_i = fitted mean, V = variance function