

FORMULARY – RISK MODELS

Review of mathematical statistics

$$f_{Y_r}(y) = \frac{n!}{(r-1)!(n-r)!} (F_X(y))^{r-1} (1-F_X(y))^{n-r} f_X(y)$$

$$(M-m) \left(2f(m)\sqrt{n} \right) \overset{\circ}{\sim} n(0;1)$$

Normal populations

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1); \quad T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{(n-1)}; \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \quad \text{where}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Large samples

$$Z = \frac{\bar{X} - \mu}{\sqrt{\text{var}(\bar{X})}} \overset{\circ}{\sim} N(0,1); \quad Z = \frac{\bar{X} - \mu}{\sqrt{\hat{\text{var}}(\bar{X})}} \overset{\circ}{\sim} N(0,1)$$

Complete data

$$S_n(x) = r_j / n, \quad y_{j-1} \leq x < y_j, \quad j = 2, 3, \dots, k$$

$$S_n(x) = 1 - \frac{c_j F_n(c_{j-1}) - c_{j-1} F_n(c_j)}{c_j - c_{j-1}} - \frac{F_n(c_j) - F_n(c_{j-1})}{c_j - c_{j-1}} x, \quad c_{j-1} \leq x < c_j$$

$$H(x) = \int_{-\infty}^x h(y) dy \quad S(x) = e^{-H(x)}$$

$$\hat{H}(x) = \sum_{i=1}^{j-1} (s_i / r_i), \quad y_{j-1} \leq x < y_j, \quad j = 2, 3, \dots, k$$

Modified Data

$$S_n(x) = \prod_{i=1}^{j-1} \left(\frac{r_i - s_i}{r_i} \right), \quad y_{j-1} \leq x < y_j, \quad j = 2, 3, \dots, k$$

$$\text{Greenwood's formula: } \hat{\text{var}}(S_n(x)) \approx S_n(x)^2 \times \sum_{i: y_i \leq x} \frac{s_i}{r_i(r_i - s_i)}$$

$$U = \exp \left(z_{\alpha/2} \frac{\sqrt{\hat{\text{var}}(S_n(x))}}{S_n(x) \times \ln S_n(x)} \right) \quad \left((S_n(x))^{UU}; (S_n(x))^U \right)$$

$$\hat{H}(x) = \sum_{i=1}^{j-1} (s_i / r_i), \quad y_{j-1} \leq x < y_j, \quad j = 2, 3, \dots, k$$

$$\hat{\text{var}}(\hat{H}(x)) = \sum_{i: y_i \leq x} (s_i / r_i^2)$$

Kernel estimation: $\hat{f}(x) = \sum_{j=1}^k p(y_j) k_{y_j}(x)$

○ Uniform kernel: $k_y(x) = 1/(2b) \quad y-b \leq x \leq y+b$

○ Triangular kernel: $k_y(x) = \begin{cases} (x-y+b)/b^2 & y-b \leq x \leq y \\ (y+b-x)/b^2 & y \leq x \leq y+b \end{cases}$

○ Gamma kernel: $k_y(x) = \frac{x^{\alpha-1} e^{-x\alpha/y}}{(y/\alpha)^\alpha \Gamma(\alpha)} I_{(0;+\infty)}(x)$

Parameter estimation

Smoothed percentiles: $\hat{\pi}_g = (1-h)x_{(j)} + hx_{(j+1)}$ where $j = \lfloor (n+1)g \rfloor$ and

$$h = (n+1)g - j$$

Maximum likelihood

(Mild regularity conditions)

$\hat{\theta}$ maximum likelihood estimator of θ . $\hat{\theta} \overset{\circ}{\sim}$ normal with mean θ and variance $I(\theta)^{-1}$.

$$I(\theta) = -n E \left(\frac{\partial^2}{\partial \theta^2} \ln f(X | \theta) \right) = -E(\ell''(\theta | X_1, X_2, \dots, X_n)); \quad I(\theta) \approx I(\hat{\theta}) \approx -H(\hat{\theta});$$

If there is more than one parameter the maximum likelihood estimators will follow an asymptotic multidimensional normal distribution. $I(\theta)$ is now a matrix with (r,s) element given by

$$I(\theta)_{r,s} = -E \left(\frac{\partial^2}{\partial \theta_r \partial \theta_s} \ell(\theta | X_1, X_2, \dots, X_n) \right)$$

Delta method

One parameter: $\hat{\theta} \overset{\circ}{\sim} n(\theta, \sigma^2/n)$ then $g(\hat{\theta}) \overset{\circ}{\sim}$ normal with mean $g(\theta)$ and variance $g'(\theta)^2 \times (\sigma^2/n)$.

k parameters: $\hat{\theta} \overset{\circ}{\sim}$ normal with mean θ and covariance matrix (Σ/n) then

$g(\theta) = g(\theta_1, \theta_2, \dots, \theta_k) \overset{\circ}{\sim}$ normal with mean $g(\theta)$ and variance $(\partial g)^T \Sigma (\partial g) / n$

Joint confidence interval

$$\{\theta : \ell(\theta) \geq c\} \text{ where } c = \ell(\hat{\theta}) - 0.5 \times q_\alpha \text{ and } q_\alpha \text{ quantile of a } \chi^2_{(r)}$$

Bayesian estimation

$$\pi_{\theta|\mathbf{X}}(\theta | \mathbf{x}) \propto L(\theta | \mathbf{x}) \times \pi(\theta)$$

$$f_{Y|\mathbf{X}}(y | \mathbf{x}) = \int f_{Y|\Theta}(y | \theta) \pi(\theta | \mathbf{x}) d\theta \quad \text{or} \quad f_{Y|\mathbf{X}}(y | \mathbf{x}) = \sum_{\theta} f_{Y|\Theta}(y | \theta) \pi(\theta | \mathbf{x})$$

Under regulatory conditions, $\theta | \mathbf{x} \overset{\circ}{\sim}$ normal

Model selection

Kolmogorov-Smirnov $D_n = \max_{j=1,2,\dots,k} (d_j)$ where

$$d_j = \max(F_n(x_j) - F^*(x_j); F^*(x_j^-) - F_n(x_j^-))$$

α	0.10	0.05	0.01
Aprox. crit. value	$1.22 / \sqrt{n}$	$1.36 / \sqrt{n}$	$1.63 / \sqrt{n}$

Anderson-Darling

$$A^2 = -n F^*(u) + n \sum_{j=0}^k (1 - F_n(y_j))^2 (\ln(1 - F^*(y_j)) - \ln(1 - F^*(y_{j+1}))) + n \sum_{j=1}^k (F_n(y_j))^2 (\ln F^*(y_{j+1}) - \ln F^*(y_j))$$

No ties and no censoring: $A^2 = -n - \sum_{j=1}^n \frac{2j-1}{n} (\ln F^*(y_j) + \ln(1 - F^*(y_{n+1-j})))$

α	0.10	0.05	0.01
Aprox. crit. value	1.933	2.492	3.857

Chi-squared

$$\chi^2 = \sum_{j=1}^k \frac{(O_j - E_j)^2}{E_j}$$

Likelihood ratio

$$\lambda(x_1, x_2, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta | x_1, x_2, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta | x_1, x_2, \dots, x_n)} \text{ and } -2 \ln \lambda(X_1, X_2, \dots, X_n) \overset{\circ}{\sim} \chi_{(r)}^2$$

Simulation

Box-Muller formulae: (U_1, U_2) independent uniform variables, then

$Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$ and $Z_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$ are independent $n(0;1)$ random variables