1/1

Lecture (10)

• 4.2.3 Finite Mixture Distributions p. (52)

Why we use mixture models ?

One motivation for mixing is that the underlying phenomenon may actually be several phenomena that occur with unknown probabilities. For example a randomly selected dental claim may be from a checkup, from a filling, from a repair (such as a crown) or from a surgical procedure. Because of the differing modes for these possibilities, a mixture model may work well.

Defn (1)

Consider a set of random variables $\{X_1, X_2, \dots, X_K\}$ with associated marginal distribution functions $\{F_{X_1}, F_{X_2}, \dots, F_{X_K}\}$. the random Variable Y is a k-point mixture of the random variables $\{X_1, X_2, \dots, X_K\}$ if its cdf is given by

$$F_Y(y) = a_1 F_{X_1}(y) + a_2 F_{X_2}(y) + \dots + a_K F_{X_K}(y),$$

where all $a_j > 0$ and $a_1 + a_2 + \dots + a_K = 1$

Example 4.4 p. (53)

For a general liability insurance models, Actuaries consider a mixture of two Pareto distributions with 4 parameters, its cdf is given by

$$F(x) = 1 - a\underbrace{\left(\frac{\theta_1}{\theta_1 + x}\right)^{\alpha}}_{\text{For large infrequent claims}} - (1-a)\underbrace{\left(\frac{\theta_2}{\theta_2 + x}\right)^{\alpha+2}}_{\text{For small frequent claims}}$$

Defn (2)

A variable-component mixture distribution has a distribution function that can be written as

$$F(x) = \sum_{j=1}^{K} a_j F_j(x), \quad \sum_{j=1}^{K} a_j = 1, \quad a_j > 0,$$

$$j = 1, \dots, K, \; K = 1, 2, \dots$$

2

Note

If each component of a mixture distn has the same parametric distribution (but different parameters), then it's called a "Variable mixture of g's" distribution, where g stands for the name of the component distribution.

• Example 4.5 p. 53

Determine the distribution, density, and hazard rate functions for the variable mixture-of-exponentials distribution.

Ans: A mixture of exponential distribution functions can be written as $F(x) = 1 - a_1 e^{-x/\theta_1} - a_2 e^{-x/\theta_2} - \cdots - a_k e^{-x/\theta_k}$,

$$\sum_{j=1}^{k} a_j = 1, \quad a_j, \theta_j > 0, \; j = 1, \ldots, k, \; k = 1, 2, \ldots$$

∴ the p.d.f $f(x)$ for the variable mixture of exponentials distn is given by

$$f(x) = a_1 \theta_1^{-1} e^{-x/\theta_1} + a_2 \theta_2^{-1} e^{-x/\theta_2} + \cdots + a_k \theta_k^{-1} e^{-x/\theta_k}$$

and hazard rate fn for this variable mixture is given by

$$h(x) = \frac{f(x)}{S(x)}$$

$$h(x) = \frac{a_1 \theta_1^{-1} e^{-x/\theta_1} + a_2 \theta_2^{-1} e^{-x/\theta_2} + \cdots + a_k \theta_k^{-1} e^{-x/\theta_k}}{a_1 e^{-x/\theta_1} + a_2 e^{-x/\theta_2} + \cdots + a_k e^{-x/\theta_k}}$$

Notes:

(1) when $k = 2$, there are three parameters $(a_1, \theta_1, \theta_2)$, where a_2 can be determined as $a_2 = 1 - a_1$.

(2) when $k = 3$, we have 5 parameters $(a_1, a_2, \theta_1, \theta_2, \theta_3)$, $a_3 = 1 - a_1 - a_2$

(3) " $k = 4$, " " 7 parameters $(a_1, a_2, a_3, \theta_1, \theta_2, \theta_3, \theta_4)$, where $a_4 = 1 - a_1 - a_2 - a_3$

— — —

Example 4.6 p. 54

Illustrate how a two-point mixture of gamma variables can create a bimodal distribution

Ans:

Consider a 50-50 mixture of two gamma distributions. One has parameters $\alpha=4$ and $\theta=7$ (for a mode of 21) and the other has parameters $\alpha=15$ and $\theta=7$ (for a mode of 98). The density function of this mixture is given by

$$f(x)=0.5\frac{(x/7)^4 e^{-x/7}}{x\,\Gamma(4)}+0.5\frac{(x/7)^{15} e^{-x/7}}{x\,\Gamma(15)}$$

$$f(x)=0.5\frac{x^3 e^{-x/7}}{3!\,7^4}+0.5\frac{x^{14} e^{-x/7}}{14!\,7^{15}}$$

For $X\sim$ gamma

$$f(x)=\frac{(x/\theta)^\alpha e^{-x/\theta}}{x\,\Gamma(\alpha)}$$

Remember

see, its graph p. 54 textbook

pb 4.7 p. 57

Seventy-five percent of claims have a normal distribution with a mean of 3000 and a variance of 1,000,000. The remaining 25% have a normal distribution with a mean of 4,000 and a variance of 1,000,000. Determine the probability that a randomly selected claim exceeds 5,000.

Ans:

For this mixture distribution,

$$F(5000)=0.75\,\Phi\left(\frac{5000-3000}{1000}\right)+0.25\,\Phi\left(\frac{5000-4000}{1000}\right)$$

$$F(5{,}000)=0.75\,\Phi(2)+0.25\,\Phi(1)$$
$$=0.75(0.9772)+0.25(0.8413)=0.9432$$

$pr(X>5{,}000)$? , X is the claim r.v

$$Pr(X>5{,}000)=1-0.9432=0.0568$$