Nonlife Actuarial Models

Chapter 12

Parametric Model Estimation

Learning Objectives

1. Methods of moments and percentile matching

2. Maximum likelihood estimation

3. Bayesian estimation

4. Cox's proportional hazards model

5. Modeling joint distributions using copula

12.1 Methods of Moments and Percentile Matching

- Let $f(\cdot;\theta)$ be the pdf or pf of a failure-time or loss variable X, where $\theta = (\theta_1, \cdots, \theta_k)'$ is a k-element parameter vector.

- We denote μ'_r as the rth raw moment of X. Assuming the functional dependence of μ'_r on θ, we write $\mu'_r(\theta)$.

- Given a random sample $\boldsymbol{x} = (x_1, \cdots, x_n)$ of X, the sample analogues of $\mu'_r(\theta)$, denoted by $\hat{\mu}'_r$, is

$$\hat{\mu}'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r. \tag{12.1}$$

12.1.1 Method of Moments

- The **method-of-moments** estimate $\hat{\theta}$ is the solution of θ in the equations
$$\mu'_r(\theta) = \hat{\mu}'_r, \qquad \text{for } r = 1, \cdots, k. \tag{12.2}$$

- Thus, we have a set of k equations involving k unknowns $\theta_1, \cdots, \theta_k$.

- We assume that a solution to the equations in (12.2) exists.

Example 12.1: Let X be the claim-frequency random variable. Determine the method-of-moments estimates of the parameter of the distribution of X, if X is distributed as (a) $\mathcal{PN}(\lambda)$, (b) $\mathcal{GM}(\theta)$, and (c) $\mathcal{BN}(m, \theta)$, where m is a known constant.

Solution: All the distributions in this example are discrete with a single parameter in the pf. Hence, $k = 1$ and we need to match only the population mean $\mathrm{E}(X)$ to the sample mean $\bar{x}$. For (a), $\mathrm{E}(X) = \lambda$. Hence, $\hat{\lambda} = \bar{x}$. For (b), we have

$$\mathrm{E}(X) = \frac{1-\theta}{\theta} = \bar{x},$$

so that

$$\hat{\theta} = \frac{1}{1+\bar{x}}.$$

For (c), we equate $\mathrm{E}(X) = m\theta$ to $\bar{x}$ and obtain

$$\hat{\theta} = \frac{\bar{x}}{m},$$

which is the sample proportion. $\square$

Example 12.2: Let X be the claim-severity random variable. Determine the method-of-moments estimates of the parameters of the distribution of X, if X is distributed as (a) $\mathcal{G}(\alpha, \beta)$, (b) $\mathcal{P}(\alpha, \gamma)$, and (c) $\mathcal{U}(a, b)$.

Solution: All the distributions in this example are continuous with 2 parameters in the pdf. Thus, $k = 2$, and we need to match the first 2 population moments μ_1' and μ_2' to the sample moments $\hat{\mu}_1'$ and $\hat{\mu}_2'$. For (a), we have

$$\mu_1' = \alpha\beta = \hat{\mu}_1' \qquad \text{and} \qquad \mu_2' = \alpha\beta^2 + \alpha^2\beta^2 = \hat{\mu}_2',$$

from which we obtain

$$\beta\mu_1' + \mu_1'^2 = \mu_2'.$$

Hence, the method-of-moments estimates are

$$\hat{\beta} = \frac{\hat{\mu}_2' - \hat{\mu}_1'^2}{\hat{\mu}_1'}$$

and

$$\hat{\alpha} = \frac{\hat{\mu}_1'}{\hat{\beta}} = \frac{\hat{\mu}_1'^2}{\hat{\mu}_2' - \hat{\mu}_1'^2}.$$

For (b), the population moments are

$$\mu_1' = \frac{\gamma}{\alpha - 1} \qquad \text{and} \qquad \mu_2' = \frac{2\gamma^2}{(\alpha - 1)(\alpha - 2)},$$

from which we obtain

$$\mu_2' = \frac{2\mu_1'^2(\alpha - 1)}{\alpha - 2}.$$

Hence,

$$\hat{\alpha} = \frac{2(\hat{\mu}_2' - \hat{\mu}_1'^2)}{\hat{\mu}_2' - 2\hat{\mu}_1'^2}$$

and

$$\hat{\gamma} = (\hat{\alpha} - 1)\hat{\mu}_1'.$$

Note that if $\hat{\mu}_2' - 2\hat{\mu}_1'^2 < 0$, then $\hat{\alpha} < 0$ and the model $\mathcal{P}(\hat{\alpha}, \hat{\gamma})$ is not well defined.

For (c), the population moments are

$$\mu_1' = \frac{a+b}{2} \qquad \text{and} \qquad \mu_2' = \frac{(b-a)^2}{12} + \mu_1'^2.$$

Solving for a and b, and evaluating the solutions at $\hat{\mu}_1'$ and $\hat{\mu}_2'$, we obtain

$$\hat{a} = \hat{\mu}_1' - \sqrt{3(\hat{\mu}_2' - \hat{\mu}_1'^2)} \qquad \text{and} \qquad \hat{b} = \hat{\mu}_1' + \sqrt{3(\hat{\mu}_2' - \hat{\mu}_1'^2)}.$$

However, if $\min\{x_1, \cdots, x_n\} < \hat{a}$, or $\max\{x_1, \cdots, x_n\} > \hat{b}$, the model $\mathcal{U}(\hat{a}, \hat{b})$ is incompatible with the claim-severity data. $\square$

- As can be seen from Example 12.2, the estimates may be incompatible with the model assumption.

- However, provided the parameters of the distribution can be solved uniquely from the population moments, the method-of-moments estimates are *consistent* for the model parameters.

- The method of moments can also be applied to censored or truncated distributions.

Example 12.3: A random sample of 15 ground-up losses, X, with a policy limit of 15 has the following observations

$$2,\ 3,\ 4,\ 5,\ 8,\ 8,\ 9,\ 10,\ 11,\ 11,\ 12,\ 12,\ 15,\ 15,\ 15.$$

If X is distributed as $\mathcal{U}(0, b)$, determine the method-of-moments estimate of b.

Solution: To estimate b we match the sample mean of the loss payments to the mean of the censored uniform distribution. The mean of the sample of 15 observations is 9.3333. As

$$\mathrm{E}\left[(X \wedge u)\right] = \int_0^u [1 - F(x)]\, dx = \int_0^u \frac{b - x}{b}\, dx = u - \frac{u^2}{2b},$$

and $u = 15$, we have

$$15 - \frac{(15)^2}{2\hat{b}} = 9.3333,$$

so that $\hat{b} = 19.8528$. $\square$

12.1.2 Method of Percentile Matching

- Some statistical distributions with thick tails (such as the **Cauchy** distribution and some members of the **stable distribution family**), do not have any moments.

- For such distributions, the method of moments breaks down.

- On the other hand, as quantiles or percentiles of a distribution always exist, we may estimate the model parameters by matching the population percentiles (as functions of the parameters) to the sample percentiles. This approach is called **the method of percentile or quantile matching**.

- Consider k quantities $0 < \delta_1, \cdots, \delta_k < 1$, and let $\delta_i = F(x_{\delta_i}; \theta)$ so that $x_{\delta_i} = F^{-1}(\delta_i; \theta)$, where θ is a k-element vector of the parameters of the df.

- We write $x_{\delta_i}(\theta)$, emphasizing its dependence on θ.

- Let $\hat{x}_{\delta_i}$ be the δ_i-quantile computed from the sample. The quantile-matching method solves for the value of $\hat{\theta}$, so that

$$x_{\delta_i}(\hat{\theta}) = \hat{x}_{\delta_i}, \qquad \text{for } i = 1, \cdots, k. \tag{12.5}$$

- Again we assume that a solution of $\hat{\theta}$ exists for the above equations, and it is called the **percentile- or quantile-matching estimate**.

Example 12.5: Let X be distributed as $\mathcal{W}(\alpha, \lambda)$. Determine the quantile-matching estimates of α and λ.

Solution: Let $0 < \delta_1, \delta_2 < 1$. From equation (2.36), we have

$$\delta_i = 1 - \exp\left[-\left(\frac{x_{\delta_i}}{\lambda}\right)^{\alpha}\right], \qquad i = 1, 2,$$

so that

$$-\left(\frac{x_{\delta_i}}{\lambda}\right)^{\alpha} = \log(1 - \delta_i), \qquad i = 1, 2.$$

We take the ratio of the case of $i = 1$ to $i = 2$ to obtain

$$\left(\frac{x_{\delta_1}}{x_{\delta_2}}\right)^{\alpha} = \frac{\log(1 - \delta_1)}{\log(1 - \delta_2)},$$

which implies

$$\hat{\alpha} = \frac{\log\left[\dfrac{\log(1 - \delta_1)}{\log(1 - \delta_2)}\right]}{\log\left(\dfrac{\hat{x}_{\delta_1}}{\hat{x}_{\delta_2}}\right)},$$

where $\hat{x}_{\delta_1}$ and $\hat{x}_{\delta_2}$ are sample quantiles. Given $\hat{\alpha}$ we further solve for $\hat{\lambda}$ to obtain

$$\hat{\lambda} = \frac{\hat{x}_{\delta_1}}{[-\log(1-\delta_1)]^{\frac{1}{\hat{\alpha}}}} = \frac{\hat{x}_{\delta_2}}{[-\log(1-\delta_2)]^{\frac{1}{\hat{\alpha}}}}.$$

Thus, analytical solutions of $\hat{\alpha}$ and $\hat{\lambda}$ are obtainable. $\square$

12.2 Bayesian Estimation Method

- The Bayesian estimator of Θ is a decision rule of assigning a value to Θ based on the observed data.
- Given a loss function, the decision rule is chosen to give as small an expected loss as possible.
- If the squared-error loss (or quadratic loss) function is adopted, the Bayesian estimator (the decision rule) is the mean of the posterior distribution (given the data) of Θ.
- See Section 8.1 for more details.

12.3 Maximum Likelihood Estimation Method

- Suppose we have a random sample of n observations of X, denoted by $\boldsymbol{x} = (x_1, \cdots, x_n)$. Given the pdf or pf of X, $f(\cdot; \theta)$, we define the **likelihood function** of the sample as the product of $f(x_i; \theta)$, denoted by $L(\theta; \boldsymbol{x})$. Thus, we have

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta), \qquad (12.6)$$

  which is taken as a function of θ given $\boldsymbol{x}$.

- As the observations are independent, $L(\theta; \boldsymbol{x})$ is the joint probability or joint density of the observations.

- We further define the **log-likelihood function** as the logarithm of

$L(\theta; \boldsymbol{x})$, i.e.,

$$\log L(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \log f(x_i; \theta). \tag{12.7}$$

- The value of θ, denoted by $\hat{\theta}$, that maximizes the likelihood function is called the **maximum likelihood estimator (MLE)** of θ.

- $\hat{\theta}$ also maximizes the log-likelihood function.

- Maximization of the log-likelihood function is often easier than maximization of the likelihood function, as the former is the *sum* of n terms involving θ while the latter is a product.

- We now discuss the asymptotic properties of the MLE and its applications.

- We first consider the case where $\boldsymbol{X}$ are independently and identically distributed. This is the case where we have complete individual loss observations.

- We then extend the discussion to the case where $\boldsymbol{X}$ are not identically distributed, such as for grouped or incomplete data.

- The properties of the MLE are well established in the statistics literature and their validity depends on some technical conditions, referred to as **regularity conditions**.

- We summarize the properties of the MLE here, with the details deferred to the Appendix A.18.

- We first consider the case where θ is a scalar. The **Fisher infor-**

mation in a single observation, denoted by $I(\theta)$, is defined as

$$I(\theta) = \mathrm{E}\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right], \tag{12.8}$$

which is also equal to

$$\mathrm{E}\left[-\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]. \tag{12.9}$$

- In addition, the **Fisher information in a random sample $\boldsymbol{X}$**, denoted by $I_n(\theta)$, is defined as

$$I_n(\theta) = \mathrm{E}\left[\left(\frac{\partial \log L(\theta;\boldsymbol{X})}{\partial\theta}\right)^2\right], \tag{12.10}$$

which is n times the Fisher information in a single observation, i.e.,

$$I_n(\theta) = nI(\theta). \tag{12.11}$$

- Also, $I_n(\theta)$ can be computed as

$$I_n(\theta) = \mathrm{E}\left[-\frac{\partial^2 \log L(\theta; \boldsymbol{X})}{\partial \theta^2}\right]. \tag{12.12}$$

- For any unbiased estimator $\tilde{\theta}$ of θ, the **Cramér-Rao inequality** states that

$$\mathrm{Var}(\tilde{\theta}) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}, \tag{12.13}$$

and an unbiased estimator is said to be **efficient** if it attains the **Cramér-Rao lower bound**.

- The MLE $\hat{\theta}$ is formally defined as

$$\hat{\theta} = \max_{\theta} \{L(\theta; \boldsymbol{x})\} = \max_{\theta} \{\log L(\theta; \boldsymbol{x})\}, \tag{12.14}$$

which can be computed by solving the first-order condition

$$\frac{\partial \log L(\theta; \boldsymbol{x})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log f(x_i; \theta)}{\partial \theta} = 0. \tag{12.15}$$

Theorem 12.1: Under certain regularity conditions, the distribution of $\sqrt{n}(\hat{\theta} - \theta)$ converges to the normal distribution with mean 0 and variance $1/I(\theta)$, i.e.,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right), \tag{12.16}$$

where $\xrightarrow{D}$ denotes convergence in distribution.

- Theorem 12.1 has several important implications.
- First, $\hat{\theta}$ is asymptotically unbiased and consistent.
- Second, in large samples $\hat{\theta}$ is approximately normally distributed with mean θ and variance $1/I_n(\theta)$.
- Third, since the variance of $\hat{\theta}$ converges to the Cramér-Rao lower bound, $\hat{\theta}$ is *asymptotically* efficient.

- We now generalize the results to the case where $\theta = (\theta_1, \cdots, \theta_k)'$ is a k-element vector. The **Fisher information matrix** in an observation is now defined as the $k \times k$ matrix

$$I(\theta) = \mathrm{E}\left[\frac{\partial \log f(X;\theta)}{\partial \theta}\frac{\partial \log f(X;\theta)}{\partial \theta'}\right], \tag{12.17}$$

which is also equal to

$$\mathrm{E}\left[-\frac{\partial^2 \log f(X;\theta)}{\partial \theta\, \partial \theta'}\right]. \tag{12.18}$$

- The Fisher information matrix in a random sample of n observations is $I_n(\theta) = nI(\theta)$.

- Let $\tilde{\theta}$ be any unbiased estimator of θ. We denote the variance matrix of $\tilde{\theta}$ by $\mathrm{Var}(\tilde{\theta})$.

- Hence, the ith diagonal element of $\mathrm{Var}(\tilde{\theta})$ is $\mathrm{Var}(\tilde{\theta}_i)$, and its (i, j)th element is $\mathrm{Cov}(\tilde{\theta}_i, \tilde{\theta}_j)$.

- Denoting $I_n^{-1}(\theta)$ as the inverse of $I_n(\theta)$, the multivariate version of the Cramér-Rao inequality states that

$$\mathrm{Var}(\tilde{\theta}) - I_n^{-1}(\theta) \tag{12.19}$$

  is a non-negative definite matrix.

- As a property of non-negative definite matrices, the diagonal elements of $\mathrm{Var}(\tilde{\theta}) - I_n^{-1}(\theta)$ are non-negative, i.e., the lower bound of $\mathrm{Var}(\tilde{\theta}_i)$ is the ith diagonal element of $I_n^{-1}(\theta)$.

- An unbiased estimator is said to be efficient if it attains the Cramér-Rao lower bound $I_n^{-1}(\theta)$.

Theorem 12.2: Under certain regularity conditions, the distribution of $\sqrt{n}(\hat{\theta} - \theta)$ converges to the **multivariate normal distribution** with mean vector 0 and variance matrix $I^{-1}(\theta)$, i.e.,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}\left(0, I^{-1}(\theta)\right). \tag{12.20}$$

- Theorem 12.2 says that the MLE is asymptotically unbiased, consistent, asymptotically normal and efficient.

- The MLE has the convenient property that it satisfies the **invariance principle**. Suppose $g(\cdot)$ is a one-to-one function and $\hat{\theta}$ is the MLE of θ, then the invariance principle states that $g(\hat{\theta})$ is the MLE of $g(\theta)$.

12.3.1 Complete Individual Data

- Complete individual observations form a random sample, for which the likelihood and log-likelihood functions are given in equations (12.6) and (12.7), respectively.
- Maximization through equation (12.15) then applies.

Example 12.8: Determine the MLE of the following models with a random sample of n observations: (a) $\mathcal{PN}(\lambda)$, (b) $\mathcal{GM}(\theta)$, (c) $\mathcal{E}(\lambda)$, and (d) $\mathcal{U}(0, \theta)$.

Solution: Note that (a) and (b) are discrete models, while (c) and (d) are continuous. The same method, however, applies. For (a) the log-likelihood function is

$$\log L(\lambda; \boldsymbol{x}) = n\bar{x} \log \lambda - n\lambda - \sum_{i=1}^{n} \log(x_i!),$$

and the first-order condition is

$$\frac{\partial \log L(\lambda; \boldsymbol{x})}{\partial \lambda} = \frac{n\bar{x}}{\lambda} - n = 0.$$

Thus, the MLE of λ is

$$\hat{\lambda} = \bar{x},$$

which is equal to the method-of-moments estimate derived in Example 12.1.

For (b) the log-likelihood function is

$$\log L(\theta; \boldsymbol{x}) = n \log \theta + [\log(1-\theta)] \sum_{i=1}^{n} x_i,$$

and the first-order condition is

$$\frac{\partial \log L(\theta; \boldsymbol{x})}{\partial \theta} = \frac{n}{\theta} - \frac{n\bar{x}}{1-\theta} = 0.$$

Solving for the above, we obtain

$$\hat{\theta} = \frac{1}{1+\bar{x}},$$

which is also the method-of-moments estimate derived in Example 12.1.

For (c) the log-likelihood function is

$$\log L(\lambda; \boldsymbol{x}) = n \log \lambda - n\lambda\bar{x},$$

with the first-order condition being

$$\frac{\partial \log L(\lambda; \boldsymbol{x})}{\partial \lambda} = \frac{n}{\lambda} - n\bar{x} = 0.$$

Thus, the MLE of λ is

$$\hat{\lambda} = \frac{1}{\bar{x}}.$$

For (d), it is more convenient to consider the likelihood function, which is

$$L(\theta; \boldsymbol{x}) = \left(\frac{1}{\theta}\right)^n,$$

for $0 < x_1, \cdots, x_n \leq \theta$, and 0 otherwise. Thus, the value of θ that maximizes the above expression is $\hat{\theta} = \max\{x_1, \cdots, x_n\}$. Note that in this case the MLE is not solved from equation (12.15).

A remark for the $\mathcal{U}(0, \theta)$ case is of interest. Note that from Theorem 12.1, we conclude that $\mathrm{Var}(\sqrt{n}\hat{\theta})$ converges to a positive constant when n tends to infinity, where $\hat{\theta}$ is the MLE. From Example 10.2, however, we learn that the variance of $\max\{x_1, \cdots, x_n\}$ is

$$\frac{n\theta^2}{(n+2)(n+1)^2},$$

so that $\mathrm{Var}(n \max\{x_1, \cdots, x_n\})$ converges to a positive constant when n

tends to infinity. Hence, Theorem 12.1 breaks down. This is due to the violation of a regularity condition for this model.

Example 12.10: Determine the asymptotic distribution of the MLE of the following models with a random sample of n observations: (a) $\mathcal{PN}(\lambda)$, and (b) $\mathcal{GM}(\theta)$. Hence, derive $100(1-\alpha)\%$ confidence interval estimates for the parameters of the models.

Solution: For (a) the second derivative of the log-likelihood of an observation is

$$\frac{\partial^2 \log f(x;\lambda)}{\partial \lambda^2} = -\frac{x}{\lambda^2}.$$

Thus,

$$I(\lambda) = \mathrm{E}\left[-\frac{\partial^2 \log f(X;\lambda)}{\partial \lambda^2}\right] = \frac{1}{\lambda},$$

so that

$$\sqrt{n}(\hat{\lambda}-\lambda) \xrightarrow{D} \mathcal{N}(0, \lambda).$$

As in Example 12.8, $\hat{\lambda} = \bar{x}$, which is also the estimate for the variance. Hence, in large samples $\bar{x}$ is approximately normally distributed with mean λ and variance λ/n (estimated by $\bar{x}/n$). A $100(1-\alpha)\%$ confidence interval of λ is computed as

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{x}}{n}}.$$

Note that we can also estimate the $100(1-\alpha)\%$ confidence interval of λ by

$$\bar{x} \pm z_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}},$$

where s^2 is the sample variance. This estimate, however, will not be as efficient if X is Poisson.

For (b) the second derivative of the log-likelihood of an observation is

$$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} = -\frac{1}{\theta^2} - \frac{x}{(1-\theta)^2}.$$

As $\mathrm{E}(X) = (1-\theta)/\theta$, we have

$$I(\theta) = \mathrm{E}\left[-\frac{\partial^2 \log f(X;\theta)}{\partial \theta^2}\right] = \frac{1}{\theta^2} + \frac{1}{\theta(1-\theta)} = \frac{1}{\theta^2(1-\theta)}.$$

Thus,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}\left(0, \theta^2(1-\theta)\right),$$

where, from Example 12.8,

$$\hat{\theta} = \frac{1}{1+\bar{x}}.$$

A $100(1-\alpha)\%$ confidence interval of θ can be computed as

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\theta}^2(1-\hat{\theta})}{n}}.$$

12.3.2 Grouped and Incomplete Data

- When the sample data are grouped and/or incomplete, the observations are no longer iid.

- Nonetheless, we can still formulate the likelihood function and compute the MLE.

- The first step is to write down the likelihood function or log-likelihood function of the sample that is appropriate for the way the observations are sampled.

- We first consider the case where we have complete observations that are grouped into k intervals: $(c_0, c_1]$, $(c_1, c_2], \cdots, (c_{k-1}, c_k]$, where $0 \leq c_0 < c_1 < \cdots < c_k = \infty$.

- Let the number of observations in the interval $(c_{j-1}, c_j]$ be n_j so that $\sum_{j=1}^{k} n_j = n$. Given a parametric df $F(\cdot;\theta)$, the probability of a single observation falling inside the interval $(c_{j-1}, c_j]$ is $F(c_j;\theta) - F(c_{j-1};\theta)$.

- Assuming the *individual* observations are iid, the likelihood of having n_j observations in the interval $(c_{j-1}, c_j]$, for $j = 1, \cdots, k$, is

$$L(\theta; \boldsymbol{n}) = \prod_{j=1}^{k} \left[F(c_j;\theta) - F(c_{j-1};\theta)\right]^{n_j}, \tag{12.21}$$

  where $\boldsymbol{n} = (n_1, \cdots, n_k)$.

- The log-likelihood function of the sample is

$$\log L(\theta; \boldsymbol{n}) = \sum_{j=1}^{k} n_j \log\left[F(c_j;\theta) - F(c_{j-1};\theta)\right]. \tag{12.22}$$

- Now we consider the case where we have individual observations that are right censored.

- If the ground-up loss is continuous, the claim amount will have a distribution of the mixed type, described by a pf-pdf. Specifically, if there is a policy limit of u, only claims of amounts in the interval $(0, u]$ are observable. Losses of amount exceeding u are censored, so that the probability of a claim of amount u is $1 - F(u; \theta)$.

- Thus, if the claim data consist of $\boldsymbol{x} = (x_1, \cdots, x_{n_1})$, where $0 < x_1, \cdots, x_{n_1} < u$, and n_2 claims of amount u, with $n = n_1 + n_2$, then the likelihood function is given by

$$L(\theta; \boldsymbol{x}, n_2) = \left[\prod_{i=1}^{n_1} f(x_i; \theta)\right] [1 - F(u; \theta)]^{n_2}. \qquad (12.23)$$

- The log-likelihood function is

$$\log L(\theta; \boldsymbol{x}, n_2) = n_2 \log\left[1 - F(u;\theta)\right] + \sum_{i=1}^{n_1} \log f(x_i;\theta). \tag{12.24}$$

- If the insurance policy has a deductible of d, the data of claim payments are sampled from a population with truncation, i.e., only losses with amounts exceeding d are sampled.

- Thus, the pdf of the ground-up loss *observed* is

$$\frac{f(x;\theta)}{1 - F(d;\theta)}, \qquad \text{for } d < x. \tag{12.25}$$

- If we have a sample of claim data $\boldsymbol{x} = (x_1, \cdots, x_{n_1})$, then the likelihood function is given by

$$L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} \frac{f(x_i;\theta)}{1 - F(d;\theta)} = \frac{1}{\left[1 - F(d;\theta)\right]^n} \prod_{i=1}^{n} f(x_i;\theta), \qquad \text{where } d < x_1, \cdots, x_n. \tag{12.26}$$

- Thus, the log-likelihood function is

$$\log L(\theta; \boldsymbol{x}) = -n \log\left[1 - F(d; \theta)\right] + \sum_{i=1}^{n} \log f(x_i; \theta). \qquad (12.27)$$

- We denote y_i as the modified loss amount, such that $y_i = x_i - d$. Let $\boldsymbol{y} = (y_1, \cdots, y_n)$. Suppose we wish to model the distribution of the payment in a payment event, and denote the pdf of this distribution by $\tilde{f}(\cdot; \theta^*)$, then the likelihood function of $\boldsymbol{y}$ is

$$L(\theta^*; \boldsymbol{y}) = \prod_{i=1}^{n} \tilde{f}(y_i; \theta^*), \qquad \text{for } 0 < y_1, \cdots, y_n. \qquad (12.28)$$

- This model is called the **shifted model**. It captures the distribution of the loss in a payment event and may be very different from the model of the ground-up loss distribution, i.e., $\tilde{f}(\cdot)$ may differ from $f(\cdot)$.

- As the observations in general may not be iid, Theorems 12.1 and 12.2 may not apply. The asymptotic properties of the MLE beyond the iid assumption are summarized in the theorem below, which applies to a broad class of models.

Theorem 12.3: Let $\hat{\theta}$ denote the MLE of the k-element parameter θ of the likelihood function $L(\theta; \boldsymbol{x})$. Under certain regularity conditions, the distribution of $\sqrt{n}(\hat{\theta} - \theta)$ converges to the multivariate normal distribution with mean vector 0 and variance matrix $\mathcal{I}^{-1}(\theta)$, i.e.,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} \mathcal{N}\left(0, \mathcal{I}^{-1}(\theta)\right), \tag{12.29}$$

where

$$\mathcal{I}(\theta) = \lim_{n \to \infty} \mathrm{E}\left[-\frac{1}{n} \frac{\partial^2 \log L(\theta; \boldsymbol{x})}{\partial \theta \, \partial \theta'}\right]. \tag{12.30}$$

- Note that $\mathcal{I}(\theta)$ requires the evaluation of an expectation and depend on the unknown parameter θ. In practical applications it may be estimated by its sample counterpart. Once $\mathcal{I}(\theta)$ is estimated, confidence intervals of θ may be computed.

Example 12.11: Let the ground-up loss X be distributed as $\mathcal{E}(\lambda)$. Consider the following cases

(a) Claims are grouped into k intervals: $(0, c_1]$, $(c_1, c_2], \cdots, (c_{k-1}, \infty]$, with no deductible nor policy limit. Let $\boldsymbol{n} = (n_1, \cdots, n_k)$ denote the numbers of observations in the intervals.

(b) There is a policy limit of u. n_1 uncensored claims with ground-up losses $\boldsymbol{x} = (x_1, \cdots, x_{n_1})$ are available, and n_2 claims have a censored amount u.

(c) There is a deductible of d, and n claims with ground-up losses $\boldsymbol{x} = (x_1, \cdots, x_n)$ are available.

(d) Policy has a deductible of d and maximum covered loss of u. n_1 uncensored claims with ground-up losses $\boldsymbol{x} = (x_1, \cdots, x_{n_1})$ are available, and n_2 claims have a censored claim amount $u - d$. Denote $n = n_1 + n_2$.

(e) Similar to (d), but there are two blocks of policies with deductibles of d_1 and d_2 for Block 1 and Block 2, respectively. The maximum covered losses are u_1 and u_2 for Block 1 and Block 2, respectively. In Block 1 there are n_{11} uncensored claim observations and n_{12} censored claims of amount $u_1 - d_1$. In Block 2 there are n_{21} uncensored claim observations and n_{22} censored claims of amount $u_2 - d_2$.

Determine the MLE of λ in each case.

Solution: The df of $\mathcal{E}(\lambda)$ is $F(x;\lambda) = 1 - e^{-\lambda x}$. For (a), using equation (12.21), the likelihood function is (with $c_0 = 0$)

$$L(\lambda; \boldsymbol{n}) = \left[\prod_{j=1}^{k-1} \left(e^{-c_{j-1}\lambda} - e^{-c_j\lambda}\right)^{n_j}\right] \left(e^{-c_{k-1}\lambda}\right)^{n_k},$$

so that the log-likelihood function is

$$\log L(\lambda; \boldsymbol{n}) = -\, c_{k-1} n_k \lambda + \sum_{j=1}^{k-1} n_j \log\left(e^{-c_{j-1}\lambda} - e^{-c_j\lambda}\right).$$

The MLE is solved by maximizing the above expression with respect to λ, for which numerical method is required.

For (b) the likelihood function is

$$L(\lambda; \boldsymbol{x}) = \left[\prod_{i=1}^{n_1} \lambda e^{-\lambda x_i}\right] e^{-\lambda u n_2},$$

and the log-likelihood function is

$$\log L(\lambda; \boldsymbol{x}) = -\lambda u n_2 - \lambda n_1 \bar{x} + n_1 \log \lambda.$$

The first-order condition is

$$\frac{\partial \log L(\lambda; \boldsymbol{x})}{\partial \lambda} = -u n_2 - n_1 \bar{x} + \frac{n_1}{\lambda} = 0,$$

which produces the MLE

$$\hat{\lambda} = \frac{n_1}{n_1 \bar{x} + n_2 u}.$$

For (c) the likelihood function is

$$L(\lambda; \boldsymbol{x}) = \frac{1}{e^{-\lambda d n}} \left[\prod_{i=1}^{n} \lambda e^{-\lambda x_i}\right],$$

and the log-likelihood function is

$$\log L(\lambda; \boldsymbol{x}) = \lambda dn - \lambda n\bar{x} + n\log\lambda.$$

The first-order condition is

$$\frac{\partial \log L(\lambda; \boldsymbol{x})}{\partial \lambda} = nd - n\bar{x} + \frac{n}{\lambda} = 0,$$

so that the MLE is

$$\hat{\lambda} = \frac{1}{\bar{x} - d}.$$

For (d) the likelihood function is

$$L(\lambda; \boldsymbol{x}) = \frac{1}{e^{-\lambda dn}} \left[\prod_{i=1}^{n_1} \lambda e^{-\lambda x_i}\right] e^{-\lambda u n_2},$$

with log-likelihood

$$\log L(\lambda; \boldsymbol{x}) = \lambda dn - \lambda n_1\bar{x} + n_1 \log\lambda - \lambda u n_2,$$

and first-order condition

$$\frac{\partial \log L(\lambda; \boldsymbol{x})}{\partial \lambda} = nd - n_1 \bar{x} + \frac{n_1}{\lambda} - u n_2 = 0.$$

The MLE is

$$\hat{\lambda} = \frac{n_1}{n_1(\bar{x} - d) + n_2(u - d)}.$$

For (e) the log-likelihood is the sum of the two blocks of log-likelihoods given in (d). Solving for the first-order condition, we obtain the MLE as

$$\begin{aligned}
\hat{\lambda} &= \frac{n_{11} + n_{21}}{n_{11}(\bar{x}_1 - d_1) + n_{21}(\bar{x}_2 - d_2) + n_{12}(u_1 - d_1) + n_{22}(u_2 - d_2)} \\
&= \frac{\sum_{i=1}^{2} n_{i1}}{\sum_{i=1}^{2} \left[n_{i1}(\bar{x}_i - d_i) + n_{i2}(u_i - d_i) \right]}.
\end{aligned}$$

$\square$

12.4 Models with Covariates

- We have so far assumed that the failure-time or loss distributions are *homogeneous*, i.e., the same distribution applies to all insured objects.

- In practice, the future lifetime of smokers and non-smokers might differ. The accident rates of teenage drivers and middle-aged drivers might differ, etc.

- We now discuss some approaches in modeling the failure-time and loss distributions in which some attributes (called the **covariates**) of the objects affect the distributions.

- Let $S(x;\theta)$ denote the survival function of interest, called the **baseline survival function**, which applies to the distribution independent of the object's attributes.

- Suppose for the ith insured object, there is a vector of k attributes, denoted by $z_i = (z_{i1}, \cdots, z_{ik})'$, which affects the survival function.

- We denote the survival function of the ith object by $S(x; \theta, z_i)$.

12.4.1 Proportional Hazards Model

- Given the survival function $S(x; \theta)$, the hazard function $h(x; \theta)$ is defined as

$$h(x; \theta) = -\frac{d \log S(x; \theta)}{dx}, \tag{12.31}$$

from which we have

$$S(x; \theta) = \exp\left(-\int_0^x h(x; \theta)\, dx\right). \tag{12.32}$$

- We now allow the hazard function to vary with the individuals and denote it by $h(x; \theta, z_i)$. In contrast, $h(x; \theta)$, which does not vary with i, is called the **baseline hazard function**.

- A simple model can be constructed by assuming that there exists a function $m(\cdot)$, such that if we denote $m_i = m(z_i)$, then

$$h(x; \theta, z_i) = m_i \, h(x; \theta). \tag{12.33}$$

- This is called the proportional hazards model, which postulates that the hazard function of the ith individual is a multiple of the baseline hazard function, and the multiple depends on the covariate z_i.

- An important implication of the proportional hazards model is that the survival function of the ith individual is given by

$$S(x; \theta, z_i) \quad = \quad \exp\left(- \int_0^x h(x; \theta, z_i) \, dx \right)$$

$$
\begin{aligned}
&= \exp\left(-\int_0^x m_i\, h(x;\theta)\, dx\right) \\
&= \left[\exp\left(-\int_0^x h(x;\theta)\, dx\right)\right]^{m_i} \\
&= [S(x;\theta)]^{m_i}. \qquad (12.34)
\end{aligned}
$$

- For equation (12.33) to provide a well-defined hazard function, m_i must be positive for all z_i. We choose

$$
m_i = \exp(\beta' z_i), \qquad (12.35)
$$

where $\beta = (\beta_1, \cdots, \beta_k)'$ is a vector of parameters.

- Thus, an individual has the baseline hazard function if $z_i = 0$. The pdf of the ith individual can be written as

$$
f(x;\theta, z_i) = -\frac{dS(x;\theta, z_i)}{dx}
$$

$$
\begin{aligned}
&= -\frac{d\left[S(x;\theta)\right]^{m_i}}{dx} \\
&= m_i \left[S(x;\theta)\right]^{m_i-1} f(x;\theta),
\end{aligned} \tag{12.36}
$$

where $f(x;\theta) = -dS(x;\theta)/dx$ is the **baseline pdf**.

- The MLE of the full model may be quite complicated even for a simple baseline model such as the exponential.

- Furthermore, it may be desirable to separate the estimation of the parameters in the proportional hazards function, i.e., β, versus the estimation of the baseline hazard function.

- The estimation can be done in two stages. The first stage involves estimating β using the **partial likelihood method**, and the second stage involves estimating the baseline hazard function using a nonparametric method, such as the Kaplan-Meier or Nelson-Aalen estimators.

- We now explain this method using the failure-time data terminology.

- Assume the data are arranged in the order $0 < y_1 < \cdots < y_m$, where $m \leq n$. There are w_j failures at time y_j and the risk set at time y_j is r_j.

- Suppose object i fails at time y_j, the partial likelihood of object i, denoted by $L_i(\beta)$, is defined as the probability of object i failing at time y_j given that some objects fail at time y_j. Thus, we have

$$
\begin{aligned}
L_i(\beta) &= \Pr(\text{object } i \text{ fails at time } y_j \,|\, \text{some objects fail at time } y_j) \\
&= \frac{\Pr(\text{object } i \text{ fails at time } y_j)}{\Pr(\text{some objects fail at time } y_j)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{h(y_j;\theta,z_i)}{\sum_{i'\in r_j} h(y_j;\theta,z_{i'})} \\
&= \frac{m_i\, h(y_j;\theta)}{\sum_{i'\in r_j} m_{i'}\, h(y_j;\theta)} \\
&= \frac{m_i}{\sum_{i'\in r_j} m_{i'}} \\
&= \frac{\exp(\beta' z_i)}{\sum_{i'\in r_j} \exp(\beta' z_{i'})}, \qquad \text{for } i = 1, \cdots, n. \qquad (12.37)
\end{aligned}
$$

- The partial likelihood of the sample, denoted by $L(\beta)$, is defined as

$$
L(\beta) = \prod_{i=1}^{n} L_i(\beta). \qquad (12.38)
$$

- Note that only β appears in the partial likelihood function, which can be maximized to obtain the estimate of β without any assumptions about the baseline hazard function and its estimates.

Example 12.15: A proportional hazards model has two covariates $z = (z_1, z_2)'$, each taking possible values 0 and 1. We denote $z_{(1)} = (0,0)'$, $z_{(2)} = (1,0)'$, $z_{(3)} = (0,1)'$ and $z_{(4)} = (1,1)'$. The failure times observed are

$$2\ (1),\ 3\ (2),\ 4\ (3),\ 4\ (4),\ 5\ (1),\ 7\ (3),\ 8\ (1),\ 8\ (4),\ 9\ (2),\ 11\ (2),\ 11\ (2),\ 12\ (3),$$

where the index i of the covariate vector $z_{(i)}$ of the observed failures are given in parentheses. Also, an object with covariate vector $z_{(2)}$ is censored at time 6, and another object with covariate vector $z_{(4)}$ is censored at time 8. Compute the partial likelihood estimate of β.

Solution: As there are 2 covariates, we let $\beta = (\beta_1, \beta_2)'$. Next we compute the multiples of the baseline hazard function. Thus, $m_{(1)} = \exp(\beta' z_{(1)}) = 1$, $m_{(2)} = \exp(\beta' z_{(2)}) = \exp(\beta_1)$, $m_{(3)} = \exp(\beta' z_{(3)}) =$

$\exp(\beta_2)$ and $m_{(4)} = \exp(\beta' z_{(4)}) = \exp(\beta_1 + \beta_2)$. We tabulate the data and the computation of the partial likelihood in Table 12.1.

Table 12.1: Computation of the partial likelihood for Example 12.15

		Covariate	r_j of covariate $z_{(i)}$				$L_j(\beta) = \text{num}_j/\text{den}_j$	
j	y_j	vector	(1)	(2)	(3)	(4)	num_j	den_j
1	2	$z_{(1)}$	3	5	3	3	$m_{(1)}$	$3m_{(1)} + 5m_{(2)} + 3m_{(3)} + 3m_{(4)}$
2	3	$z_{(2)}$	2	5	3	3	$m_{(2)}$	$2m_{(1)} + 5m_{(2)} + 3m_{(3)} + 3m_{(4)}$
3	4	$z_{(3)}$, $z_{(4)}$	2	4	3	3	$m_{(3)}m_{(4)}$	$\left[2m_{(1)} + 4m_{(2)} + 3m_{(3)} + 3m_{(4)}\right]^2$
4	5	$z_{(1)}$	2	4	2	2	$m_{(1)}$	$2m_{(1)} + 4m_{(2)} + 2m_{(3)} + 2m_{(4)}$
5	7	$z_{(3)}$	1	3	2	2	$m_{(3)}$	$m_{(1)} + 3m_{(2)} + 2m_{(3)} + 2m_{(4)}$
6	8	$z_{(1)}$, $z_{(4)}$	1	3	1	2	$m_{(1)}m_{(4)}$	$\left[m_{(1)} + 3m_{(2)} + m_{(3)} + 2m_{(4)}\right]^2$
7	9	$z_{(2)}$	0	3	1	0	$m_{(2)}$	$3m_{(2)} + m_{(3)}$
8	11	$z_{(2)}$, $z_{(2)}$	0	2	1	0	$m^2_{(2)}$	$\left[2m_{(2)} + m_{(3)}\right]^2$
9	12	$z_{(3)}$	0	0	1	0	$m_{(3)}$	$m_{(3)}$

If two objects, i and i', have the same failure time y_j, their partial likelihoods have the same denominator (see equation (12.37)). With a slight abuse of notation, we denote $L_j(\beta)$ as the partial likelihood of the object (or the product of the partial likelihoods of the objects) with failure time y_j. Then the partial likelihood of the sample is equal to

$$L(\beta) = \prod_{i=1}^{12} L_i(\beta) = \prod_{j=1}^{9} L_j(\beta) = \prod_{j=1}^{9} \frac{\text{num}_j}{\text{den}_j},$$

where num_j and den_j are given in the last two columns of Table 12.1. Maximizing $L(\beta)$ with respect to β, we obtain $\hat{\beta}_1 = -0.6999$ and $\hat{\beta}_2 = -0.5518$. These results imply $\hat{m}_{(1)} = 1$, $\hat{m}_{(2)} = 0.4966$, $\hat{m}_{(3)} = 0.5759$ and $\hat{m}_{(4)} = 0.2860$. □

- Having estimated the parameter β in the proportional hazards model, we can continue to estimate the baseline hazard function nonparametrically using the Nelson-Aalen method.

- Equation (12.40) may be modified as follows

$$\hat{H}(y;\theta) = \sum_{\ell=1}^{j} \frac{w_\ell}{r_\ell^*}, \qquad \text{for } y_j \le y < y_{j+1}, \tag{12.42}$$

where r_ℓ^* is the modified risk set defined by

$$r_\ell^* = \sum_{i' \in r_\ell} m_{i'}. \tag{12.43}$$

Example 12.16: For the data in Example 12.15, compute the Nelson-Aalen estimate of the baseline hazard function and the baseline survival function. Estimate the survival functions $S(3.5; z_{(2)})$ and $S(8.9; z_{(4)})$.

Solution: The results are summarized in Table 12.2. Note that r_ℓ^* in Column 4 are taken from the last Column of Table 12.1 (ignore the square, if any) evaluated at $\hat{\beta}$.

Table 12.2: Nelson-Aalen estimates for Example 12.16

ℓ	y_ℓ	w_ℓ	r^*_ℓ	$\dfrac{w_\ell}{r^*_\ell}$	$\hat{H}(y)=\sum_{j=1}^{\ell}\dfrac{w_j}{r^*_j}$	$\hat{S}(y)=\exp\left[-\hat{H}(y)\right]$
1	2	1	8.0689	0.1239	0.1239	0.8834
2	3	1	7.0689	0.1414	0.2654	0.7669
3	4	2	6.5723	0.3043	0.5697	0.5656
4	5	1	5.7104	0.1751	0.7448	0.4748
5	7	1	4.2137	0.2373	0.9821	0.3745
6	8	2	3.6378	0.5497	1.5319	0.2161
7	9	1	2.0658	0.4840	2.0159	0.1331
8	11	2	1.5691	1.2745	3.2905	0.0372
9	12	1	0.5759	1.7363	5.0269	0.0065

We can now compute the survival functions for given covariates. In particular, we have

$$\hat{S}(3.5; z_{(2)}) = (0.7669)^{0.4966} = 0.8765,$$

and

$$\hat{S}(8.9; z_{(4)}) = (0.2161)^{0.2860} = 0.6452.$$

The values of $\hat{m}_{(2)} = 0.4966$ and $\hat{m}_{(4)} = 0.2860$ are taken from Example 12.15.

12.4.2 Generalized Linear Model

- A modeling strategy in which the mean of the loss variable X, denoted by μ, is assumed to be a function of the covariate z.

- To ensure the mean loss is positive, we adopt the following model

$$\mathrm{E}(X) = \mu = \exp(\beta' z). \tag{12.45}$$

- The exponential function used in the above equation is called the **link function**, which relates the mean loss to the covariate.

12.4.3 Accelerated Failure-Time Model

- In the accelerated failure-time model, the survival function of object i with covariate z_i, $S(x; \theta, z_i)$, is related to the baseline (i.e., $z = 0$) survival function as follows

$$S(x; \theta, z_i) = S(m_i x; \theta, 0), \tag{12.47}$$

where $m_i = m(z_i)$ for an appropriate function $m(\cdot)$.

- Again, a convenient assumption is $m(z_i) = \exp(\beta' z_i)$.

- We now denote $X(z_i)$ as the failure-time random variable for an object with covariate z_i. The expected lifetime (at birth) is

$$\mathrm{E}\left[X(z_i)\right] = \int_0^\infty S(x; \theta, z_i)\, dx$$

$$
\begin{aligned}
&= \int_0^\infty S(m_i x; \theta, 0)\, dx \\
&= \frac{1}{m_i} \int_0^\infty S(x; \theta, 0)\, dx \\
&= \frac{1}{m_i} \mathrm{E}\left[X(0)\right]. \qquad (12.48)
\end{aligned}
$$

- Hence, the expected lifetime at birth of an object with covariate z_i is $1/m_i$ times the expected lifetime at birth of a *baseline* object.

12.5 Modeling Joint Distribution using Copula

- In many practical applications researchers are often required to analyze multiple risks of the same group, similar risks from different groups, or different aspects of a risk group.
- Thus, techniques for modeling multivariate distributions are required.
- The use of **copula** provides a flexible approach to modeling multivariate distributions.

Definition 12.1: A bivariate copula $C(u_1, u_2)$ is a mapping from the unit square $[0, 1]^2$ to the unit interval $[0, 1]$. It is increasing in each component and satisfies the following conditions

1. $C(1, u_2) = u_2$ and $C(u_1, 1) = u_1$, for $0 \leq u_1, u_2 \leq 1$,

2. For any $0 \leq a_1 \leq b_1 \leq 1$ and $0 \leq a_2 \leq b_2 \leq 1$, $C(b_1, b_2) - C(a_1, b_2) - C(b_1, a_2) + C(a_1, a_2) \geq 0$.

- A bivariate copula is in fact a joint df on $[0, 1]^2$ with standard uniform marginals, i.e., $C(u_1, u_2) = \Pr(U_1 \leq u_1, U_2 \leq u_2)$, where U_1 and U_2 are uniformly distributed on $[0, 1]$.

- Let $F_{X_1 X_2}(\cdot, \cdot)$ be the joint df of X_1 and X_2, with marginal df $F_{X_1}(\cdot)$ and $F_{X_2}(\cdot)$. The theorem below, called the **Sklar Theorem**, states the representation of the joint df using a copula. It also shows how a joint distribution can be created via a copula.

Theorem 12.4: Given the joint and marginal df of X_1 and X_2, there exists a unique copula $C(\cdot, \cdot)$, such that

$$F_{X_1 X_2}(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2)). \tag{12.51}$$

Conversely, if $C(\cdot, \cdot)$ is a copula, and $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ are univariate df of X_1 and X_2, respectively, then $C(F_{X_1}(x_1), F_{X_2}(x_2))$ is a bivariate df with marginal df $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$.

Theorem 12.6: Let X_1 and X_2 be two continuous distributions with pdf $f_{X_1}(\cdot)$ and $f_{X_2}(\cdot)$, respectively. If the joint df of X_1 and X_2 is given by equation (12.51), their joint pdf can be written as

$$f_{X_1 X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) c(F_{X_1}(x_1), F_{X_2}(x_2)), \tag{12.54}$$

where

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \, \partial u_2} \tag{12.55}$$

is called the **copula density**.

Proof: This can be obtained by differentiating equation (12.51). □

- From Theorem 12.6, we can conclude that the log-likelihood of a bivariate random variable with df given by equation (12.51) is

$$\log\left[f_{X_1X_2}(x_1,x_2)\right] = \log\left[f_{X_1}(x_1)\right] + \log\left[f_{X_2}(x_2)\right] + \log\left[c(F_{X_1}(x_1), F_{X_2}(x_2))\right], \tag{12.56}$$

which is the log-likelihood of two *independent* observations of X_1 and X_2, plus a term which measures the dependence.

- We now introduce some simple bivariate copulas. **Clayton's copula**, denoted by $C_C(u_1, u_2)$, is defined as

$$C_C(u_1,u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-\frac{1}{\alpha}}, \qquad \alpha > 0. \tag{12.57}$$

The Clayton copula density is given by

$$c_C(u_1,u_2) = \frac{1+\alpha}{(u_1u_2)^{1+\alpha}}\left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-2-\frac{1}{\alpha}}. \tag{12.58}$$

- **Frank's copula**, denoted by $C_F(u_1, u_2)$, is defined as

$$C_F(u_1, u_2) = -\frac{1}{\alpha} \log \left[1 + \frac{\left(e^{-\alpha u_1} - 1\right)\left(e^{-\alpha u_2} - 1\right)}{e^{-\alpha} - 1} \right], \qquad \alpha \neq 0, \tag{12.59}$$

which has the following copula density

$$c_F(u_1, u_2) = \frac{\alpha e^{-\alpha(u_1+u_2)}\left(1 - e^{-\alpha}\right)}{\left[e^{-\alpha(u_1+u_2)} - e^{-\alpha u_1} - e^{-\alpha u_2} + e^{-\alpha}\right]^2}. \tag{12.60}$$

- Another popular copula is the **Gaussian copula** defined by

$$C_G(u_1, u_2) = \Psi_\alpha(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \qquad -1 < \alpha < 1, \tag{12.61}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal df and $\Psi_\alpha(\cdot, \cdot)$ is the df of a standard bivariate normal variate with correlation coefficient α. The Gaussian copula density is

$$c_G(u_1, u_2) = \frac{1}{\sqrt{1-\alpha^2}} \exp\left[-\frac{\eta_1^2 - 2\alpha\eta_1\eta_2 + \eta_2^2}{2(1-\alpha^2)}\right] \exp\left[\frac{\eta_1^2 + \eta_2^2}{2}\right], \tag{12.62}$$

where $\eta_i = \Phi^{-1}(u_i)$, for $i = 1, 2$.

Example 12.19: Let $X_1 \sim \mathcal{W}(0.5, 2)$ and $X_2 \sim \mathcal{G}(3, 2)$, and assume that Clayton's copula with parameter α fits the bivariate distribution of X_1 and X_2. Determine the probability $p = \Pr(X_1 \leq \mathrm{E}(X_1),\, X_2 \leq \mathrm{E}(X_2))$ for $\alpha = 0.001$, 1, 2, 3 and 10.

Solution: The means of X_1 and X_2 are

$$\mathrm{E}(X_1) = 2\Gamma(3) = 4 \qquad \text{and} \qquad \mathrm{E}(X_2) = (2)(3) = 6.$$

Let $u_1 = F_{X_1}(4) = 0.7569$ and $u_2 = F_{X_2}(6) = 0.5768$, so that

$$p = \Pr(X_1 \leq 4, X_2 \leq 6)$$

$$
\begin{aligned}
&= C_C(0.7569, 0.5768) \\
&= \left[(0.7569)^{-\alpha} + (0.5768)^{-\alpha} - 1\right]^{-\frac{1}{\alpha}}.
\end{aligned}
$$

The computed values of p are

α	0.001	1	2	3	10
p	0.4366	0.4867	0.5163	0.5354	0.5734

Note that when X_1 and X_2 are independent, $p = (0.7569)(0.5768) = 0.4366$, which corresponds to the case where α approaches 0. The dependence between X_1 and X_2 increases with α, as can be seen from the numerical results. $\square$