Problem 1. Inversion Theorems

Show the one-to-one relationship between the probability distribution function and the characteristic function. Express the probability distribution function as a function of the characteristic function.

Inversion Formula

There is a one-to-one correspondence between cumulative distribution functions and characteristic functions, so it is possible to find one of these functions if we know the other. The formula in the definition of the characteristic function allows us to compute ϕ when we know the distribution function F (or density f). If, on the other hand, we know the characteristic function ϕ and want to find the corresponding distribution function, then one of the following inversion theorems can be used.

Theorem: If the characteristic function ϕ_X of a random variable X is integrable, then F_X is absolutely continuous, and therefore X has a probability density function. In the univariate case (i.e., when X is scalar-valued), the density function is given by:

$$f_X(x) = F'_X(x) = \frac{1}{2\pi} \int_{\mathbf{R}} e^{-itx} \varphi_X(t) \, dt.$$

In the multivariate case, it is:

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} e^{-i(t \cdot x)} \varphi_X(t) \lambda(dt),$$

where $t \cdot x$ is the dot product.

The density function is the Radon‑Nikodym derivative of the distribution μ_X with respect to the Lebesgue measure λ:

$$f_X(x) = \frac{d\mu_X}{d\lambda}(x).$$

Problem 2.

Explain why the moment generation function does not exist when a certain moment is absent. Provide an explanation for the absence of a moment in the context of the logistic distribution.

Problem 3.

Logistic distribution 有對稱性。
説明：其他 model 沒有對稱性

As a set of independent binary regressions, to arrive at the multinomial logit model, one can imagine, for K possible outcomes, running $K-1$ independent binary logistic regression models, in which one outcome is chosen as a "pivot" and then the other $K-1$ outcomes are separately regressed against the pivot outcome. If outcome K (the last outcome) is chosen as the pivot, the $K-1$ regression equations are:

$$\ln\left(\frac{\Pr(Y_i = k)}{\Pr(Y_i = K)}\right) = \boldsymbol{\beta}_k \cdot \mathbf{X}_i, \quad k < K. \tag{1}$$

This formulation is also known as the Additive Log Ratio transform, commonly used in compositional data analysis. In other applications, it's referred to as "relative risk".

If we exponentiate both sides and solve for the probabilities, we get:

$$\Pr(Y_i = k) = \Pr(Y_i = K) \cdot e^{\boldsymbol{\beta}_k \cdot \mathbf{X}_i}, \quad k < K. \tag{2}$$

Using the fact that all K of the probabilities must sum to one, we find:

$$\Pr(Y_i = K) = 1 - \sum_{j=1}^{K-1} \Pr(Y_i = j) = 1 - \sum_{j=1}^{K-1} \Pr(Y_i = K) \cdot e^{\boldsymbol{\beta}_j \cdot \mathbf{X}_i} \Rightarrow \Pr(Y_i = K) = \frac{1}{1 + \sum_{j=1}^{K-1} e^{\boldsymbol{\beta}_j \cdot \mathbf{X}_i}}. \tag{3}$$

We can use this to find the other probabilities:

$$\Pr(Y_i = k) = \frac{e^{\boldsymbol{\beta}_k \cdot \mathbf{X}_i}}{1 + \sum_{j=1}^{K-1} e^{\boldsymbol{\beta}_j \cdot \mathbf{X}_i}}, \quad k < K. \tag{4}$$

$$\Pr(Y_i = l) = \frac{\mathbf{I}(l = K) + \mathbf{I}(l \neq K)e^{\boldsymbol{\beta}_l \cdot \mathbf{X}_i}}{1 + \sum_{j \neq K} e^{\boldsymbol{\beta}_j \cdot \mathbf{X}_i}}, l = 1, 2, \cdots, K. \tag{5}$$

The fact that we run multiple regressions reveals why the model relies on the assumption of independence of irrelevant alternatives described above.

Problem 4.

(b.1) Log-linear model:

$$P(Y = y) = C(\theta_1, \theta_2) \exp\left(Y^T \theta_1 + W^T \theta_2\right),$$

where

- $W = (Y_1 Y_2, Y_1 Y_3, \cdots, Y_{m-1} Y_m, \cdots, Y_1 Y_2 \cdots Y_m)^T$,

- $\theta_1 = \left(\theta_1^{(1)}, \cdots, \theta_m^{(1)}\right)$

- $\theta_2 = \left(\theta_{12}^{(2)}, \cdots, \theta_{m-1m}^{(2)}, \cdots, \theta_{12\cdots m}^{(m)}\right)$

- $C(\theta_1, \theta_2)$ is a function of θ_1 and θ_2 that normalizes the p.d.f. to integrate to one.

Transformation:

$$(\theta_1, \theta_2) \to (\mu, \theta_2), \mu = (\mu_1, \ldots, \mu_m) \triangleq \mu(\theta_1, \theta_2).$$

Model assumption:

$$\operatorname{logit}\left(\mu_j\right) = X_j^T \beta.$$

The score equation for β under this parameterization takes the GEE form:

$$\left(\frac{\partial \mu}{\partial \beta}\right)^T [V(Y)]^{-1} (Y - \mu) = 0,$$

where

$$\frac{\partial \mu}{\partial \beta} = \left(\frac{\partial \mu_1}{\partial \beta}, \ldots, \frac{\partial \mu_m}{\partial \beta}\right)^T.$$

Remark:

The conditional odds ratios are not easily interpreted when the association among responses is itself a focus of the study.

Properties:

1. From

$$M_Y(t) = E\left[e^{t^T Y}\right] = \sum_y C(\theta_1, \theta_2) \exp\left(y^T(t + \theta_1) + w^T \theta_2\right) = \frac{C(\theta_1, \theta_2)}{C(\theta_1 + t, \theta_2)},$$

one has

- $$\mu = \left.\frac{\partial M_Y(t)}{\partial t}\right|_{t=0} = -\frac{\frac{\partial}{\partial \theta_1} C(\theta_1, \theta_2)}{C(\theta_1, \theta_2)},$$

- $$E\left[YY^T\right] = \left.\frac{\partial^2 M_Y(t)}{\partial t \partial t^T}\right|_{t=0} = -\frac{\frac{\partial^2}{\partial \theta_1 \partial \theta_1^T} C(\theta_1, \theta_2)}{C(\theta_1, \theta_2)} + 2\mu\mu^T,$$

- $$V(Y) = -\frac{\frac{\partial^2}{\partial \theta_1 \partial \theta_1^T} C(\theta_1, \theta_2)}{C(\theta_1, \theta_2)} + \mu\mu^T.$$

2. Let

$$l(\theta_1, \theta_2) = \ln P(Y = y) = \ln\{C(\theta_1, \theta_2) + (Y^T \theta_1 + W^T \theta_2)\}.$$

We can derive that

$$\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_1} = \frac{\frac{\partial}{\partial \theta_1} C(\theta_1, \theta_2)}{C(\theta_1, \theta_2)} + Y = (Y - \mu),$$

and hence,

$$\frac{\partial l(\theta_1, \theta_2)}{\partial \beta} = \left(\frac{\partial \mu}{\partial \beta}\right)^T \frac{\partial \theta_1}{\partial \mu} \left(\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_1}\right) = \left(\frac{\partial \mu}{\partial \beta}\right)^T \left(\frac{\partial \mu}{\partial \theta_1}\right)^{-1} (Y - \mu) = \left(\frac{\partial \mu}{\partial \beta}\right)^T [V(Y)]^{-1}(Y - \mu).$$

Problem 5. Show orthonormal

Let $P_{[1]}(Y = y) = \prod_{j=1}^m \mu_j^{y_j} \left(1 - \mu_j\right)^{1-y_j}$, $g(y) = P(Y = y)/P_{[1]}(Y = y)$, and V be a vector space of real-valued functions f on Y_1 (2^m possible values of y).

Here, V is regarded as an inner-product space with

$$< f_1, f_2 > \triangleq E_{P_{[1]}}[f_1 f_2] = \sum_{y \in Y_1} f_1(y) f_2(y) P_{[1]}(y).$$

It follows easily that the set of functions $S = \{1, r_1, \ldots, r_m; r_1 r_2, \ldots, r_{m-1} r_m; \ldots, r_1 r_2 \cdots r_m\}$ on Y_1 is **orthonormal** and, thus, is a basis in V. Since $g(y)$ is a function on Y_1, there exists a unique representation as a linear combination of functions in S, namely,

$$g(y) = \sum_{f \in S} < g, f > f.$$

$$\because < g, f > = \sum_{y \in Y_1} g(y) f(y) P_{[1]}(y) = \sum_{y \in Y_1} f(y) P(Y = y) = E_P[f] \, \forall f, \text{ and}$$

$$E_P[1] = 1, E_P[r_j] = 0, E_P[r_j r_k] = \rho_{jk}, \cdots, \text{ and } E_P[r_1 \cdots r_m] = \rho_{12\cdots m}.$$

$$\therefore g(y) = \left(1 + \sum_{j<k} \rho_{jk} r_j r_k + \sum_{j<k<l} \rho_{jkl} r_j r_k r_l + \cdots + \rho_{12\cdots m} r_1 r_2 \cdots r_m\right).$$