EX1

roi hezkiyahu

28 10 2022

Q1

Question 1.

Using the factorization theorem, find the sufficient statistic for a sample from the Poisson distribution. What is the distribution of the sufficient statistic? Is it minimal? Show its completeness using the *definition* of a complete statistic.

$$\text{let } Y \sim Pois(\lambda)$$
$$L(y,\lambda) = \Pi_{i=1}^{n} \frac{\lambda^{y_i} e^{-\lambda}}{y_i!} = \frac{\lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}}{\Pi_{i=1}^{n} y_i!}$$
$$\text{let: } h(y) := \frac{1}{\Pi_{i=1}^{n} y_i!}, \quad g(\lambda, \sum_{i=1}^{n} y_i) = \lambda^{\sum_{i=1}^{n} y_i} e^{-n\lambda}$$
$$\text{from the factorization therom we can conlude that } \sum_{i=1}^{n} y_i \text{ is a sufficent statistic.}$$
$$(1) \text{ let us prove that } \sum_{i=1}^{n} y_i \sim Pois(n\lambda)$$
$$\text{let X,Y be two poisson random variable with } \lambda_x, \lambda_y \text{ respectivly, and let us show } X+Y \sim Pois(\lambda_x+\lambda_y)$$
$$P(X+Y=k) = \sum_{i=0}^{k} P(X=i)P(Y=k-i) = \sum_{i=0}^{k} \frac{\lambda_x^i e^{-\lambda_x}}{i!} \frac{\lambda_y^{(k-i)} e^{-\lambda_y}}{(k-i)!} = e^{-(\lambda_x+\lambda_y)} \sum_{i=0}^{k} \frac{\lambda_x^i \lambda_y^{k-i}}{(k-i)!i!} = \frac{e^{-(\lambda_x+\lambda_y)}}{k!} \sum_{i=0}^{k} \frac{\lambda_x^i \lambda_y^{k-i} k!}{(k-i)!i!} =$$
$$= \frac{e^{-(\lambda_x+\lambda_y)}}{k!} \sum_{i=0}^{k} \frac{(\lambda_x^i \lambda_y^{k-i})k!}{(k-i)!i!} = \frac{e^{-(\lambda_x+\lambda_y)}}{k!} \sum_{i=0}^{k} \lambda_x^i \lambda_y^{k-i} \binom{n}{k} = \frac{e^{-(\lambda_x+\lambda_y)}}{k!} (\lambda_x+\lambda_y)^k$$
$$\text{from the proof above and with induction we get (1)}$$
$$\text{lets show its minimality:}$$
$$\frac{L(y_1,\lambda)}{L(y_2,\lambda)} = \frac{\frac{\lambda^{\sum_{i=1}^{n} y_{1i}} e^{-n\lambda}}{\Pi_{i=1}^{n} y_{1i}!}}{\frac{\lambda^{\sum_{i=1}^{n} y_{2i}} e^{-n\lambda}}{\Pi_{i=1}^{n} y_{2i}!}} = \frac{\Pi_{i=1}^{n} y_{2i}!}{\Pi_{i=1}^{n} y_{1i}!} \lambda^{\sum_{i=1}^{n} y_{1i} - \sum_{i=1}^{n} y_{2i}}$$
$$\text{the liklhidood ratio is independed of } \lambda \text{ if} \sum_{i=1}^{n} y_{1i} = \sum_{i=1}^{n} y_{2i} \Rightarrow \sum_{i=1}^{n} y_i \text{ is a minial sufficient statistic}$$
$$\text{lastly lets show completness:}$$
$$E(g(\sum_{i=1}^{n} y_i)) = \sum_{i=0}^{\infty} P(\sum_{i=1}^{n} y_i = k) g(k) = \sum_{i=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} g(k) = 0 \iff g(k) = 0 \text{ thus } \sum_{i=1}^{n} y_i \text{ is complete}$$

Q2

Question 2.

Suppose we have a sample of n independent observations $Y_1, \ldots, Y_n$, where each Y_i belongs to one of K categories: P(Y belongs to the j-th category)=p_j, j=1,...,K; p_1+...+p_k=1 (multinomial distribution). Show that this distribution belongs to the exponential family with *k-1* parameters. Find the natural parameters and sufficient statistic. Is it complete? Calculate the means and the variance-covariance matrix of the sufficient statistic.

$$L(y,p) = \Pi_{i=1}^{k} p_i^{\sum_{l=1}^{n} I(y_l=i)} = \frac{n!}{\Pi_{i=1}^{k}\sum_{l=1}^{n} I(y_l=i)}(1-\sum_{m=1}^{k-1} p_m)^{n-\sum_{j=1}^{n} I(y_i<k)}\Pi_{i=1}^{k-1} p_i^{\sum_{l=1}^{n} I(y_l=i)}$$

$$\text{denote } \frac{n!}{\Pi_{i=1}^{k}\sum_{l=1}^{n} I(y_l=i)} =: e^{d(y)}$$

$$loglik(y,p) = d(y) + (n - \sum_{j=1}^{n} I(y_i < k))ln(1-\sum_{m=1}^{k-1} p_m) + \sum_{i=1}^{k-1}[ln(p_i)\sum_{l=1}^{n} I(y_l=i)] =$$

$$= d(y) + nln(1-\sum_{m=1}^{k-1} p_m) + \sum_{i=1}^{k-1}[ln\left(\frac{p_i}{1-\sum_{m=1}^{k-1} p_m}\right)\sum_{l=1}^{n} I(y_l=i)]$$

$$\text{denote: } T_j(y) = \sum_{l=1}^{n} I(y_l=j) \quad \forall i<k, \quad C_j(p) = ln\left(\frac{p_j}{1-\sum_{m=1}^{k-1} p_m}\right) \quad \forall i<k \quad S(p) = nln(1-\sum_{m=1}^{k-1} p_m)$$

$$\text{thus we get: } loglik(y,p) = S(p) + \sum_{i=1}^{k-1} T_j(y)C_j(p)$$

T has k-1 paramaters so we can conclude that the multinomial distribtuion belongs to the exponintioal family with k-1 paramaters

lets see if T is complete:

$$\text{denote } x_j = \sum_{l=1}^{n} I(y_l=j)$$

$$E(g(T_j(y))) = E(g(x_j)) = \sum_{i=0}^{n} P(x_j=i)g(x_j) = 0 \iff P(x_i=j) = 0\, \forall i \; or \; g(x_j) = 0$$

$$P(x_i=j) > 0 \; thus \; E(g(T_j(y))) = 0 \iff g(x_j) = 0$$

we get the T is complete

$$E(T_j) = E(\sum_{l=1}^{n} I(y_l=j)) = \sum_{l=1}^{n} E(I(y_l=j)) = np_j$$

$$E(T) = n(p_1,\ldots,p_{k-1})^t =: n\tilde{p}^t$$

$$V(T_j) = n(p_j)(1-p_j); \quad \text{(we know the variance for binomial distribution)}$$

$$V(T_i+T_j) = n(p_j+p_i)(1-p_j-p_i); \quad (T_i+T_j \sim B(n, p_i+p_j))$$

$$2Cov(T_j,T_i) = V(T_i+T_j) - V(T_j) - V(T_i) = n(p_j+p_i)(1-p_j-p_i) - n(p_j)(1-p_j) - n(p_i)(1-p_i) =$$

$$= np_j + np_i - np_j^2 - 2np_jp_i - np_i^2 - np_j + np_j^2 - np_i + np_i^2 = -2np_jp_i \Rightarrow Cov(T_j,T_i) = -np_jp_i$$

$$V(T) = -n(\tilde{p})\tilde{p}^t + Diag(n\tilde{p})$$

Q3

Question 3.

Let $Y_1,\ldots,Y_n$ be a sample of independent observations, where $Y_i \sim f_{\theta_i}(y)$ is from the one parameter exponential family (i.e. each Y_i belongs to the same class of distributions $f_\theta(y)$ from the exponential family but not necessarily with the same parameter), and $\eta_i = c(\theta_i)$ are the corresponding natural parameteres. The following (generalized linear regression) model describes η_i as a function of p explanatory variables $x_1,\ldots,x_p$:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

Show that the joint distribution of the data belongs to the exponential family with $p+1$ parameters, find the natural parameters and sufficient statistic. Write down the likelihood function as a function of $\boldsymbol{\beta}$ for the particular case of Poisson distribution.

$$x_0 := \underline{1}$$

$$f_{\theta_i}(y) = exp\{\eta_i T_i(y_i) + d_i(\theta_i) + S(y_i)\}$$

$$L(\theta, y) = \Pi_{i=1}^{n} f_{\theta_i}(y) = exp\{\sum_{i=1}^{n} \eta_i T_i(y_i) + nd(\theta) + \sum_{i=1}^{n} S(y_i)\} =$$

$$= exp\{\sum_{i=1}^{n}\sum_{j=0}^{p+1} \beta_j x_{ji} T_i(y_i) + nd(\theta) + \sum_{i=1}^{n} S(y_i)\} = exp\{\sum_{j=0}^{p+1} \beta_j \sum_{i=1}^{n} x_{ji} T_i(y_i) + nd(\theta) + \sum_{i=1}^{n} S(y_i)\}$$

$$C_j(\theta) = \beta_j, \quad T_j'(y) = \sum_{i=1}^{n} x_{ji} T_i(y_i)$$

thus the distribution belongs to the p+1 paramaters exponential family with the above sufficent statistic and natural paraamters

$$Y_i \sim Pois(\lambda_i)$$

we saw that a sufficent statistic in the poission regression is: $\sum_{j=1}^{n_i} Y_{ij}$

thus in the Poission distribution case:

$$L(\lambda, y) = exp\{\sum_{j=0}^{p+1} \beta_j \sum_{i=1}^{n} x_{ji} \sum_{m=1}^{n_i} Y_{im} + nd(\lambda) + \sum_{i=1}^{n} S(y_i)\}$$

where: $d(\lambda) = ln(\lambda) - n\lambda, \quad S(y_i) = ln\left(\Pi_{j=1}^{n_i} y_{ij}!\right)$

Q4

Question 4.

1. Let $X_1 \sim Pois(\lambda_1)$, $X_2 \sim Pois(\lambda_2)$ and $X_3 \sim Pois(\lambda_1 \cdot \lambda_2)$, where X_1, X_2, X_3 are independent. Does the joint distribution of (X_1, X_2, X_3) belong to an exponential family? Find the minimal sufficient statistic for (λ_1, λ_2). Is it complete?
2. Repeat the previous paragraph, where $X_3 \sim Pois(\lambda_1 + \lambda_2)$.

4.1

$$L(\lambda_1, \lambda_2, X_1, X_2, X_3) = \Pi_{k=1}^{3}(\Pi_{i=1}^{n_k} x_{ki}!)^{-1} \lambda_1^{\sum_{j=1}^{n_1} x_{1j}} e^{-n_1\lambda_1} \lambda_2^{\sum_{j=1}^{n_2} x_{2j}} e^{-n_2\lambda_2} (\lambda_2\lambda_1)^{\sum_{j=1}^{n_3} x_{3j}} e^{-n_3\lambda_2\lambda_1}$$

$$denote\ d(X) = ln(\Pi_{k=1}^{3}(\Pi_{i=1}^{n_k} x_{ki}!)^{-1})$$

$$loglik(\lambda_1, \lambda_2, X_1, X_2, X_3) = d(X) - n_1\lambda_1 - n_2\lambda_2 - n_3\lambda_1\lambda_2 + \sum_{j=1}^{n_1} x_{1j} ln(\lambda_1) + \sum_{j=1}^{n_2} x_{2j} ln(\lambda_2) + \sum_{j=1}^{n_3} x_{3j} ln(\lambda_1) + \sum_{j=1}^{n_3} x_{3j} ln(\lambda_2) =$$

$$= d(X) + S(\lambda) + (\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_3} x_{3j}) ln(\lambda_1) + (\sum_{j=1}^{n_2} x_{2j} + \sum_{j=1}^{n_3} x_{3j}) ln(\lambda_2)$$

$$denote:\ T_1(X) = (\sum_{j=1}^{n_1} x_{1j} + \sum_{j=1}^{n_3} x_{3j}), \quad T_2(X) = (\sum_{j=1}^{n_2} x_{2j} + \sum_{j=1}^{n_3} x_{3j}), \quad C_1(\lambda) = ln(\lambda_1), \quad C_2(\lambda) = ln(\lambda_2)$$

$$the\ set\ \{(ln(\lambda_1), ln(\lambda_2)); \lambda_i \in \mathbb{R}^+\}\ contains\ an\ open\ 2d\ rectangele\ thus\ the\ sufficent\ statistic\ is\ complete\ (and\ minimal)$$

thus the joint distribution of (X_1, X_2, X_3) belongs to the 2 paramater exponential family with the minimal and complete sufficent stat

$$(T_1(X), T_2(X))$$

4.2

$$L(\lambda_1, \lambda_2, X_1, X_2, X_3) = \Pi_{k=1}^{3}(\Pi_{i=1}^{n_k} x_{ki}!)^{-1} \lambda_1^{\sum_{j=1}^{n_1} x_{1j}} e^{-n_1\lambda_1} \lambda_2^{\sum_{j=1}^{n_2} x_{2j}} e^{-n_2\lambda_2} (\lambda_2 + \lambda_1)^{\sum_{j=1}^{n_3} x_{3j}} e^{-n_3(\lambda_2+\lambda_1)}$$

$$loglik(\lambda_1, \lambda_2, X_1, X_2, X_3) = d(X) - n_1\lambda_1 - n_2\lambda_2 - n_3\lambda_1 - n_3\lambda_2 + \sum_{j=1}^{n_1} x_{1j} ln(\lambda_1) + \sum_{j=1}^{n_2} x_{2j} ln(\lambda_2) + \sum_{j=1}^{n_3} x_{3j} ln(\lambda_1 + \lambda_2) =$$

$$= d(X) + S'(\lambda) + \sum_{j=1}^{n_1} x_{1j} ln(\lambda_1) + \sum_{j=1}^{n_2} x_{2j} ln(\lambda_2) + \sum_{j=1}^{n_3} x_{3j} ln(\lambda_1 + \lambda_2)$$

$$denote:\ T_1'(X) = \sum_{j=1}^{n_1} x_{1j}, \quad T_2'(X) = \sum_{j=1}^{n_2} x_{2j}, \quad T_3'(X) = \sum_{j=1}^{n_3} x_{3j}$$

$$C_1'(\lambda) = ln(\lambda_1), \quad C_2'(\lambda) = ln(\lambda_2), \quad C_3'(\lambda) = ln(\lambda_1 + \lambda_2)$$

thus the joint distribution of (X_1, X_2, X_3) belongs to the 3 paramater exponential family with the sufficent statistic $(T_1(X), T_2(X), T_3($

$$\frac{L(\lambda_1,\lambda_2,X_1,X_2,X_3)}{L(\lambda_1,\lambda_2,X'_1,X'_2,X'_3)} = \frac{\Pi_{k=1}^{3}(\Pi_{i=1}^{n_k}x_{ki}!)^{-1}\lambda_1^{\sum_{j=1}^{n_1}x_{1j}}e^{-n_1\lambda_1}\lambda_2^{\sum_{j=1}^{n_2}x_{2j}}e^{-n_2\lambda_2}(\lambda_2+\lambda_1)^{\sum_{j=1}^{n_3}x_{3j}}e^{-n_3(\lambda_2+\lambda_1)}}{\Pi_{k=1}^{3}(\Pi_{i=1}^{n_k}x'_{ki}!)^{-1}\lambda_1^{\sum_{j=1}^{n_1}x'_{1j}}e^{-n_1\lambda_1}\lambda_2^{\sum_{j=1}^{n_2}x'_{2j}}e^{-n_2\lambda_2}(\lambda_2+\lambda_1)^{\sum_{j=1}^{n_3}x'_{3j}}e^{-n_3(\lambda_2+\lambda_1)}}$$

the liklihood ratio is independet of λ if:

$$\sum_{j=1}^{n_1}x_{1j} = \sum_{j=1}^{n_1}x'_{1j}, \quad \sum_{j=1}^{n_2}x_{2j} = \sum_{j=1}^{n_2}x'_{2j}, \quad \sum_{j=1}^{n_3}x'_{3j} = \sum_{j=1}^{n_3}x'_{3j}$$

thus the sufficent statistic we derived above is minimal

$$E(g(T_i(X))) = \sum_{k=0}^{\infty}P(T_i(X)=k)g(k) = 0 \iff g(k)=0 \Rightarrow \text{T is complete}$$