UDC: 519.224
Conference Paper

Distribution theory for exponential families

Bedrije Bexheti [1], Teuta Jusufi-Zenku [1], Qamile Asani [1]

[1] *Department of Mathematics, Faculty of Natural Sciences and Mathematics, UT*
[*] *Corresponding Author: e-mail: bedrije.bexheti@unite.edu.mk*

Abstract

The importance of exponential family distributions lies in the way in which the parameter of the model interacts with the argument of the density or frequency function in the model function. This type of structure simplifies certain aspects of the distribution theory of the model, particularly those aspects concerned with how the distributions change under changes in parameter values. Consider a random variable *Y* with model function of the form $\exp\{c(\theta)^T T(y) - A(\theta)\}h(y);$ only the first of these terms depends on θ and that term depends on *Y* only thought T(Y). In this paper we give two theorems which give expressions of the idea that the dependence of the distribution of *Y* on θ is primarily through the dependence of distribution of *T(Y)* on θ.

Keywords: random variable, natural parameter, conditional distribution, continuous distribution.

1. Introduction

Distribution theory lies at the interface of probability and statistics. It is closely related to probability theory; however, it differs in its focus on the calculation and approximation of probability distributions. Distribution theory plays a central role in the development of statistical methodology; distribution theory itself does not deal with issues of statistical inference.

In the part below we gave some definitions and theorems that will help to proof the results on this paper.
Consider random variables X and Y. From elementary probability theory we know that the conditional probability that X ∈ A given that Y = y is given by

$$\Pr(X \in A|Y = y) = \frac{\Pr(X \in A, Y = y)}{\Pr(Y = y)}$$

Since $\Pr(X \in A|Y = y)$ defines a probability distribution for X, for each y, there exists a distribution function $F_{X|Y}(x|y)$ such that

$$\Pr(X \in A|Y = y) = \int_A dF_{X|Y}(x|y);$$

the distribution function $F_{X|Y}(\cdot|y)$ is called the *conditional distribution* function of X given Y = y. If the conditional distribution of X given Y = y is absolutely continuous, then

$$\Pr(X \in A|Y = y) = \int_A p_{X|Y}(x|y)dx$$

where $p_{X|Y}(\cdot|y)$ denotes the *conditional density* of X given Y = y. If the conditional distribution of X given Y = y is discrete, then

$$\Pr(X \in A | Y = y) = \sum_{x\in A} p_{X|Y}(x|y)$$

where $p_{X|Y}(\cdot|y)$ denotes the *conditional frequency function* of X given Y = y.

Consider a random vector of the form (X, Y), where each of X and Y may be a vector and suppose that the range of (X, Y) is of the form $\mathrm{X}_1 \times \mathrm{Y}_1$ so that $X \in \mathrm{X}_1$ and $Y \in \mathrm{Y}_1$. The probability distribution of X, given by

$$\Pr(X \in A) = \Pr(X \in A, Y \in Y_1), A \subset \mathrm{X}_1$$

is called the *marginal distribution* of X .

Consider a family $\wp$ of probability distributions. A parameterization of $\wp$ is a mapping from a parameter space Θ to the set $\wp$ so that $\wp$ may be represented $\wp = \{P(\cdot;\theta); \theta \in \Theta\}$. Hence, corresponding to any statement regarding the elements P of $\wp$ is an equivalent statement regarding the elements θ of Θ.

Many frequently used families of distributions have a common structure. Consider a family of distributions on $\mathbf{R}^d$, $\{P(\cdot\,;\theta) : \theta \in \Theta\}$, such that each distribution in the family is either absolutely continuous or discrete with support not depending on θ. For each θ, let $P(\cdot\,;\theta)$ denote either the density function or frequency function corresponding to $P(\cdot\,;\theta)$. The family

of distributions is said to be an *m-parameter exponential family* if each $p(\cdot\,;\theta)$ may be written

$$p(y;\theta) = \exp\{c(\theta)^T T(y) - A(\theta)\}h(y), \quad \mathrm{y} \in \mathrm{Y} \tag{1.1}$$

where $\mathrm{Y} \subset \mathrm{R}^d, \mathrm{c} : \Theta \to \mathrm{R}^m, \mathrm{T} : \mathrm{Y} \to \mathrm{R}^m, A : \Theta \to R$ and $\mathrm{h} : \mathrm{Y} \to \mathrm{R}^+$

It is important to note that the representation (1.1) is not unique; for example, we may replace $c(\theta)$ by $c(\theta)/2$ and $T(y)$ by $2T(y)$.

It is often convenient to re-parameterize the models in order to simplify the structure of the exponential family representation. For instance, consider the reparameterization $\eta = c(\theta)$ so that the model function (1.1) becomes

$$\exp\{\eta^T T(y) - A[\theta(\eta)]\}h(y), \quad \mathrm{y} \in \mathrm{Y}.$$

Writing $k(\eta)$ for $A[\theta(\eta)]$, the model function has the form

$$\exp\{\eta^T T(y) - k(\eta)\}h(y), \quad \mathrm{y} \in \mathrm{Y}. \tag{1.2}$$

the parameter space of the model is given by

$$\mathrm{H}_0 = \{\eta \in R^m ; \eta = c(\theta), \theta \in \Theta\}.$$

The model function (1.2) is called the *canonical form* of the model function and the parameter η is called the *natural parameter* of the exponential family distribution. Note that the function k can be obtained from T, h and Y. For instance, if the distribution is absolutely continuous, we must have

$$\int_y \exp\{\eta^T T(y) - k(\eta)\}h(y)dy = 1, \ \eta \in \mathrm{H}_0$$

so that

$$k(\eta) = \log \int_y \exp\{\eta^T T(y)\}h(y)dy.$$

Theorem 1.1: Let (X,Y) denote a random vector with range $X \times Y$ and let g denote a real-valued function on X.

(i) Suppose that $E\left[\left|g(x)\right|\right] < \infty$ and let Z denote a real-valued function of Y such that $E\left[\left|Z\right|\right] < \infty$. If

$$Z = E\left[\left|g(x)\right|Y\right|\right] \text{ with probability 1}$$

then

$$E\left[Zh(Y)\right] = E\left[g(X)h(Y)\right] \qquad (2.1)$$

for all functions $h: Y \to R$ such that

$$E\left[\left|h(Y)\right|\right] < \infty \text{ and } E\left[\left|g(X)h(Y)\right|\right] < \infty$$

(ii) If Z is a function of Y such that (2.1) holds for all bounded functions $h: Y \to R$ then

$$Z = E\left[g(x)\middle|Y\right] \text{ with probability 1}$$

Theorem 1.2: Let (X, Y) denote a random vector with range $X \times Y$ and distribution function F. Let F_X and F_Y denote the marginal distribution functions of X and Y, respectively.
(i) X and Y are independent if and only if for all x, y

$$F(x, y) = F_X(x)F_Y(y)$$

(ii) X and Y are independent if and only if for all bounded functions $g_1 : X \to R$ and $g_2 : Y \to R$

$$E\left[g_1(X)g_2(Y)\right] = E\left[g_1(X)\right]E\left[g_2(Y)\right]$$

Definition: Let X denote a real-valued random variable and suppose there exists a number δ > 0 such that E[exp{t X}] < ∞ for |t| < δ. In this case, we say that X has *moment-generating function*

$$M(t) \equiv MX\ (t) = E[\exp\{t\ X\}],\ |t| < \delta;$$

δ is known as the *radius of convergence* of M_x .

Corollary 1. Let X denote a random vector taking values in R^d and let $X = (X_1, X_2)$ where X_1 takes values in R^{d_1} and X_2 takes values in R^{d_2} . Let M denote the moment-generating function of X with radius of convergence δ, let M_1 denote the moment-generating function of X_1 with radius of convergence δ_1, and let M_2 denote the moment-generating function of X_2 with radius of convergence δ_2 .

X_1 and X_2 are independent if and only if there exists a $\delta_0 > 0$ such that for all

$t = (t_1, t_2), t_1 \in R^{d_1}, t_2 \in R^{d_2}, \left\|t\right\| < \delta_{0,}$

$$M(t) = M_1(t_1)M(t_2)$$

If $p(y;\theta)$ is of the form (1.1) and θ_0 and θ_1 are two elements of the parameter space, then $\log\left[p(y;\theta_1)/p(y;\theta_0)\right]$ is a linear function of T(y) with coefficients depending on θ_0, θ_1 :

$$\log\frac{p(y;\theta_1)}{p(y;\theta_0)} = A(\theta_0) - A(\theta_1) + \left[c(\theta_1) - c(\theta_0)\right]^T T(y).$$

This type of structure simplifies certain aspects of the distribution theory of the model, particularly those aspects concerned with how the distributions change under changes in parameter values. The following lemma gives a relationship between expectations under two different parameter values.

Lemma 1: Let Y denote a random variable with model function of the form

$$\exp\left\{\eta^T T(y) - k(\eta)\right\}h(y), \quad \text{y} \in \text{Y}.$$

where $\eta \in H.$

Fix $\eta_0 \in H$ and let $g : \mathrm{Y} \to \mathrm{R}$. Then

$$E[g(Y);\eta] = \exp\{k(\eta_0) - k(\eta)\}E[g(y)\exp\{(\eta-\eta_0)^T T(y)\};\eta_0]$$

for any $\eta \in H$ such that

$$E[|g(Y)|;\eta] < \infty.$$

Consider a random variable *Y* with model function of the form

$$\exp\{c(\theta)^T T(y) - A(\theta)\}h(y);$$

this function can be written as the product of two terms, the term given by the exponential function and *h(y).* Note that only the first of these terms depends on θ and that term depends on *y* only thought T(Y). This suggest that, in some sense, the dependence of the distribution of Y on θ is primarily through the dependence of the distribution of T(Y) on θ. The following two theorems give some expressions of this idea.

2. Main results

Theorem 2.1. Let Y denote a random variable with model function of the form

$$\exp\{c(\theta)^T T(y) - A(\theta)\}h(y), y \in Y,$$

where $\theta \in \Theta.$ Then the conditional distribution of *Y* given *T(Y)* does not depend on θ.

Proof: Let $\eta = c(\theta), H$ denote the natural parameter space of the model, and

$$H_0 = \{\eta \in H : \eta = c(\theta), \theta \in \Theta\}.$$

Then the model function for this model can be written

$$\exp\{\eta^T T(y) - k(\eta)\}h(y), y \in Y$$

where $\eta \in H_0$. Hence, it suffices to show that the conditional distribution of Y given T(y), based on this model, with the parameter space enlarged to H, does not depend on η.

We proved this result by showing that for any bounded , real-valued function g on Y, $E[g(Y)|T;\eta]$does not depend on η.

Fix $\eta_0 \in H$. The idea of the proof is that the random variable

$$Z = E[g(y)|T;\eta_0]$$

satisfies

$$E[Zh(T);\eta] = E[g(Y)h(T);\eta]$$

for any $\eta \in H$, for all bounded functions h of T. Hence, by Theorem 1,

$$Z = E[g(Y)|T;\eta].$$

That is, for all $\eta_0, \eta \in H$,

$$E[g(Y)|T;\eta] = E[g(Y)|T;\eta_0],$$

which proves the result.

We now consider the details of the argument. Let h denote a bounded, real- valued function on the range of T. Then, since Z and g(Y) are bounded,

$$E[|Zh(T)|;\eta] < \infty \text{ and } E[|g(Y)h(T)|;\eta] < \infty;$$

By Lemma 1,

$$E[Zh(T);\eta] = \exp\{k(\eta) - k(\eta_0)\}E[Zh(T)\exp\{(\eta-\eta_0)^T T\};\eta_0]$$

and

$$E[g(Y)h(T);\eta] = \exp\{k(\eta) - k(\eta_0)\} E[g(Y)h(T)\exp\{(\eta - \eta_0)^T T\};\eta_0].$$

Let

$$h_0(T) = h(T)\exp\{(\eta - \eta_0)^T T\}.$$

Note that

$$E[|h_0(T)|;\eta_0] = \exp\{k(\eta_0) - k(\eta)\} E[|h(T)|;\eta] < \infty$$

It follows that

$$E[Zh_0(T);\eta_0] = E[g(Y)h_0(T);\eta_0]$$

so that

$$E[Zh(T);\eta] = E[g(Y)h(T);\eta]$$

for all bounded h. Hence, by Theorem 1.1,

$$Z \equiv E[g(Y)|T;\eta_0] = E[g(Y)|T;\eta],$$

proving the result.

Theorem 2.2: Let Y denote a random variable with model function of the form

$$\exp\{c(\theta)^T T(y) - A(\eta)\}h(y), y \in Y,$$

where $\theta \in \Theta$ and $c : \Theta \to R^m$. Let

$$H_0 = \{\eta \in H; \eta = c(\theta), \theta \in \Theta\},$$

where H denote the natural parameter space of the exponential family, and let Z denote a real-valued function on Y.

(i) If Z and T(Y) are independent, then the distribution of Z does not depend on $\theta \in \Theta$.

(ii) If H_0 contains an open subset of R^m and the distribution of Z does not depend on $\theta \in \Theta$, then Z and T(Y) are independent.

Proof. We begin by reparameterizing the model in terms of the natural paramet1er $\eta = (\theta)$ so that the model function can be written

$$\exp\{\eta^T T(y) - k(\eta)\}h(y), y \in Y,$$

with parameter space H_0.

Suppose that Z and T(Y) are independent. Define

$$\varphi(t;\eta) = E[\exp(itZ);\eta], t \in R, \eta \in H,$$

Then, by Lemma 1, for any $\eta_0 \in H$

$$\varphi(t;\eta) = \exp\{k(\eta_0) - k(\eta)\} E[\exp(itZ)\exp\{(\eta - \eta_0)^T T(Y)\};\eta_0], t \in R, \eta \in H,$$

Since Z and T(Y) are independent,

$$\varphi(t;\eta) = \exp\{k(\eta_0) - k(\eta)\} E[\exp(itZ) : \eta_0] E[\exp\{(\eta - \eta_0)^T T(Y)\};\eta_0] t \in R, \eta \in H,$$

Since

$$E[\exp\{(\eta - \eta_0)^T T(Y)\};\eta_0] = \exp\{k(\eta) - k(\eta_0)\}$$

and $E[\exp(itZ) : \eta_0] = \varphi(t;\eta_0)$, it follows that, for all $\eta, \eta_0 \in H$,

$$\varphi(t;\eta) = \varphi(t;\eta_0), t \in R$$

so that the distribution of Z does not depend on $\eta \in H$ and, hence, it does not depend on $\eta \in H_0$. This proves (i).

Now suppose that the distribution of Z does not depend on $\eta \in H_0$ and that there exists a subset H_0, H_1, such that H_1 is an open subset of R^m. Fix $\eta_0 \in H_1$ and let g denote a bounded function on the range of Z; then there exists a $\delta_1 > 0$ such that $\exp[tg(Z)]$ is bounded for $|t| < \delta_1$. Hence, by Lemma 1, for any $\eta \in H$,

$$E\{\exp[tg(Z)];\eta\} = \exp\{k(\eta_0) - k(\eta)\}E[\exp[(tg(Z)]\exp\{(\eta-\eta_0)^T T(Y)\};\eta_0]$$

Using the fact that

$$E[\exp\{(\eta-\eta_0)^T T(Y)\};\eta_0] = \exp\{k(\eta) - k(\eta_0)\},$$

it follows that, for all $\eta \in H$ and all $|t| < \delta_1$,

$$E\{\exp[tg(Z) + (\eta-\eta_0)^T T(Y)];\eta_0\} = E\{\exp[(tg(Z)];\eta\}E[\exp\{(\eta-\eta_0)^T T(Y)\};\eta_0]$$

For $\delta > 0$, let

$$H(\delta) = \{\eta \in H : \|\eta - \eta_0\| < \delta\}$$

and let $\delta_2 > 0$ be such that $H(\delta_2) \subset H_1$; since H_1 open subset of R^m and $\eta_0 \in H_1$, such a δ_2 must exist. Then , since the distribution of g(Z) does not depend on η for $\eta \in H_0$, for $\eta \in H(\delta_2)$ and $|t| < \delta_1$,

$$E\{\exp[tg(Z)];\eta\} = E\{\exp[tg(Z)];\eta_0\}.$$

It follows that, for all $\eta \in H(\delta_2)$ and all $|t| < \delta_1$,

$$E\{\exp[tg(Z) + (\eta-\eta_0)^T T(Y)];\eta_0\} = E\{\exp[(tg(Z)];\eta_0\}E[\exp\{(\eta-\eta_0)^T T(Y)\};\eta_0]$$

That is, the joint moment-generating function of g(Z) and T(Y) can be factored into the product of the two marginal moment- generating functions. Hence, by Corollary 1 g(Z) and T(Y) are independent and by part (ii) of Theorem 1.2, Z and T(Y) are independent, proving part (ii) of the theorem.

Conclusions

The importance of exponential family distributions lies in the way in which the parameter of the model interacts with the argument of the density or frequency function in the model function. This type of structure simplifies certain aspects of the distribution theory of the model, particularly those aspects concerned with how the distributions change under changes in parameter values.

References

[1]. Billingsley, P. Convergence of Probability Measures. *Wiley*, New York, 1968.

[2]. Brown, L.D. Fundamentals of Statistical Exponential Families. *IMS Lecture Notes Monograph Series 9*. IMS, Hayward, 1988.

[3]. Carl N. Morris. 1982. Natural Exponential Families with Quadratic Variance Functions. *The Annals of Statistics*, Vol. 10, No. 1,pp. 65-80.

[4]. Karr, A.F. Probability. *Springer-Verlag*, New York, 1993.

[5]. Martin J. Wainwright and Michael I. Jordan. 2008. Graphical Models, Exponential Families and Variational Inference, *Foundations and Trends in Machine Learning*, Vol. 1: No. 1–2, pp. 1-305.

[6]. Thomas A. Severini. Elements of Distribution Theory. *Cambridge University Press. USA: New York ,* 2005.

[7]. Llukan Puka. Probabilitetet dhe Statistika e Zbatuar: Konceptet themelore. *SHBLS e Re,* Tiranë. 2008.