**GRADUATION PROJECT**

**REPORT**

**GOES TO GENERAL RECORD**

**DATA**

İsmet Artuç

090120516

**Project Supervisor**

**Doç. Dr. Mustafa NADAR**

**Faculty of Science and Lectures, Mathematical Engineering**

**Istanbul Technical University**

**1. Introduction**

The ordered random variables play important roles in the theory and practice of statistics. They possess significant statistical properties. Over the last few decades, many articles on various topics of ordered statistical data have appeared. It was a special interest to coordinate and edit an interesting research problem based on material contributed by several important researchers from all over the world. In this study the estimation of parameters with different methods using the kth record values from Burr XII distribution will be discussed and based on different estimation hazard function will be obtained. At the beginning we will give brief definition of kth record data, hazard function and Burr XII distribution function.

Order statistics and record values appear in many statistical applications and are widely used in statistical modeling and inference. A form of the joint distribution of *n* ordered random variables is presented that enables a unified approach to a variety of models of ordered random variables, e.g. order statistics and record values. Generalized order statistics, provide a suitable approach to explain similarities and analogies in the two models and to generalize related results. The definition of *random kth record* can be shown as below.

Let *X*1*, X*2*, . . .* be a sequence of independent and identically distributed (i.i.d.) random variables with continuous distribution function *F* (*x*) = *P* (*X*1 ≤ *x*). Denote by *X*1,n≤ *. . .* ≤ *Xn,n* the order statistics of *X*1*, . . . , Xn*. For a fixed integer *k* ≥ 1, we define the corresponding *k*th *record times*, {*L*(*n, k*)*, n* ≥ 1}, and *k*th *record values* , {*X*(*n, k*)*, n* ≥ 1}, by setting

*L*(1*, k*) = *k, L*(*n* + 1*, k*) = min{*j > L*(*n, k*) : *Xj > Xj*−*k,j*−1} for *n* ≥ 1*,* and

*X*(*n, k*) = *XL*(*n,k*)−*k*+1*,L*(*n,k*) for *n* ≥ 1*.*

Let *N* be a positive integer-valued random variable which is independent of the *Xi*. The random variables *X*(*N, k*) are called the *random kth record.*

Now we will discuss the hazard function. Hazard function is generally used when calculating the age of an electronic device or any material. The failure rate of a system usually depends on time, with the rate varying over the life cycle of the system. For example, an automobile's failure rate in its fifth year of service may be many times greater than its failure rate during its first year of service.

The hazard function can be defined as

h(t)=\frac{f(t)}{1-F(t)}=\frac{f(t)}{R(t)}.

Many probability distributions can be used to model the failure distribution.

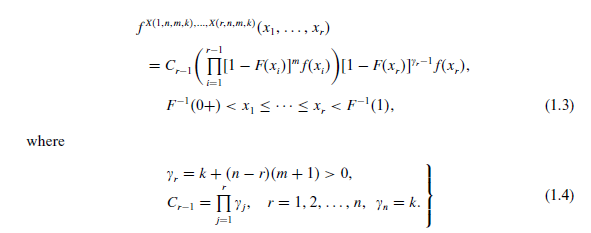
Burr distribution was first introduced by Burr (1942) as a two-parameter family. An additional scale parameter was introduced by Tadikamalla (1980). The Burr distribution can fit a wide range of empirical data. Different values of its parameters cover a broad set of skewness and kurtosis. Hence, it is used in various fields such as finance, hydrology, and reliability to model a variety of data types. Examples of data modeled by the Burr distribution are household income, crop prices, insurance risk, travel time, flood levels, and failure data.

Hazard functions of Burr type XII distribution is,



Kamps (1995) introduced the concept of generalized order statistics (gos) as a unified approach to order statistics, record values, and sequential order statistics. The gos are defined using quintile transformation based on the distribution function F.

Let X(1, n, m,k),X(r, n,m,k), k > 1, m is a real number, be gos based on absolutely continuous distribution function F with density function f. The joint density function of the above quantities is given by



with *n* ∈ *N* , *k>* 0 and *m* ∈ *R*. For more details of gos, see Kamps (1995). In the case *m* = 0 and *k* = 1, *X(r, n, m, k)* reduces to the ordinary r-th order statistics and (1.3) is the joint pdf of *r* ordinary order statistics, *X*1*:n* ≤ *X*2*:n* ≤ ··· ≤ *Xr:n*. For various distributional properties of ordinary order statistics, see David (1981) and Arnold et al. (1992). If *m* = −1 and *k* = 1, then (1.3) is the joint pdf of *r* upper record values. For some distributional properties of record values, see Ahsanullah (1995) and Arnold et al. (1998).

Distribution properties of gos from a uniform distribution are given by Ahsanullah (1996). He obtained the minimum variance linear unbiased estimators of the parameters of the two parameters of uniform distribution based on the ﬁrst *m* gos Kamps (1996) characterized the uniform distribution based on distribution properties of subranges of gos. Kamps and Gather (1997) characterized the exponential distributions by distributional properties of gos Cramer and Kamps (1996, 1998, 2001) studied some estimation problems with different sequential k-out-of-n systems. Ahsanullah (2000) gave some distributional properties of the gos from the two parameter exponential distribution. He also obtained the minimum variance linear unbiased estimators of the two parameters and characterized the exponential distribution based on gos. Cramer and Kamps (2000) derived relations for expectations of functions of gos from a class of distributions which includes the exponential, uniform, Pareto, Lomax and, Pearson I. Habibullah and Ahsanullah (2000) obtained estimates for the parameters of the Pareto distribution based on gos. Kamps and Cramer (2001) studied some distribution properties of the gos from the Pareto, power and Weibull distributions. Jaheen (2002) considered the prediction of future gos from a general class of distributions which includes the Weibull, compound Weibull, Burr type XII, Pareto, beta, and Gompertz by using Bayesian two-sample prediction technique.

**2. Maximum Likelihood Estimation**

The log likelihood function dropping terms that do not involve  and , is

 (22)

We assume that the parameters  and  are unknown. To obtain the normal equation for the unknown parameters, we differentiate (22) partially with respect to  and  ,and equate to zero, the resulting equations are

 (23)

and

 (24)

where  and .

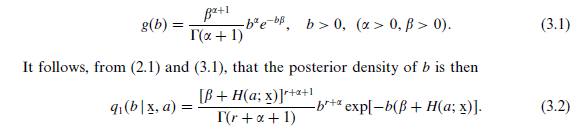
The solution of the above equation are the maximum likelihood estimators of the Burr XII () parameters and , denoted  and , respectively. As the equation expressed in (23) and (24) con not be solved analytically, one must use a numerical procedure to solve them.

**3. Bayes Estimation**

In the following, Bayesian estimation for the parameters of the Burr type XII distribution is considered for two cases, the first is when the parameter a is known and the second is when the two parameters a and b are assumed to be unknown.

**3.1.** *One Parameter Case (a is Known)*

When the parameter a is assumed to be known, we use the gamma conjugate prior density for the parameter b, that was first used by Papadopoulos (1978) and AL-Hussaini et al. (1992), when a was assumed known, in the following form

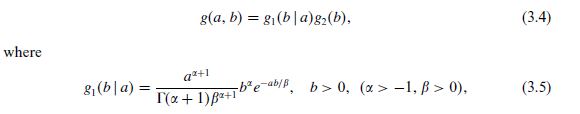


Under a squared error loss function, the Bayes estimate of the parameter b is the posterior mean in the form



**3.2.** *Two Parameter Case (a and b are Unknown)*

When both of the two parameters a and b are unknown, AL-Hussaini and Jaheen (-992) suggested a bivariate prior density function, given by



is the gamma conjugate prior, that was first used by Papadopoulos (1978), when a was assumed known, and



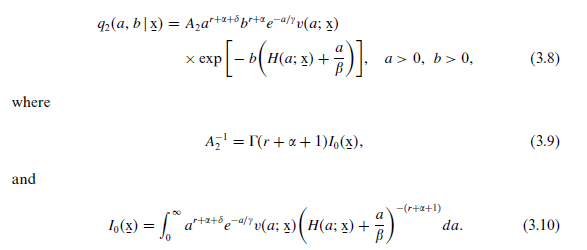
is the gamma density function.

Multiplying *g1(b* | *a)* by *g2(a),* we obtain the bivariate density of a and *b,* given from (3.4), by

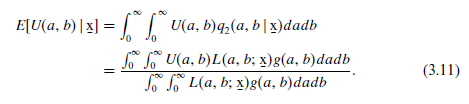


where a > -1, *f,* y, and *5* are positive real numbers and A-1 = r(£)T(a + *1)y5fa+1.*

The four-parameter gamma-gamma prior (3.7) is so chosen that it would be rich enough to cover the prior belief of the experimenter (AL-Hussaini and Jaheen, 1992). It follows, from (2.1) and (3.7), that the joint posterior density function of *a* and b given the data is thus



It is well known that, under squared error loss function, the Bayes estimator of a function, say *U(a, b),* is the posterior mean of the function and is given by a ratio of two integrals which may be written as



Generally, the ratio of two integrals (3.11) cannot be obtained in a simple closed form. Numerical methods of integration may be used in this case, which can be computationally intensive, especially in high dimensional parameter space. Instead, we can use the approximation form due to Lindley (1980). In the following, a review of Lindley’s approximation form is given.

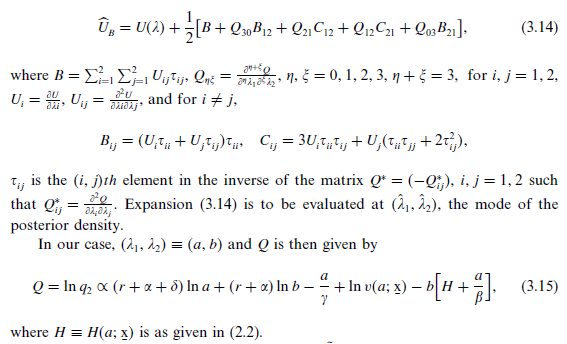
*3.2.1. The Approximation form of Lindley.* Lindley (1980) developed approximate procedures for the evaluation of the ratio of two integrals in the form



where *k = (k1,...,kN), L(k)* is the logarithm of the likelihood function, g(A) and *m(k) = U(k)g(k)* are arbitrary functions of k. From Eq. (3.12), the posterior expectation of the function *U(k),* for given x, is



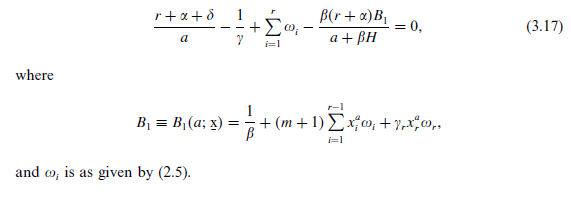
where *Q(k) = L(k) + p(k)* is the logarithm of the posterior distribution of k except for the normalizing constant and *p(k) =* ln *g(k).* Expanding *Q(k)* in (3.13) into a Taylor series expansion about the posterior mode of *k*, Lindley obtained the required expression for *E[U(k)* | x]. For more details, see Lindley (1980).

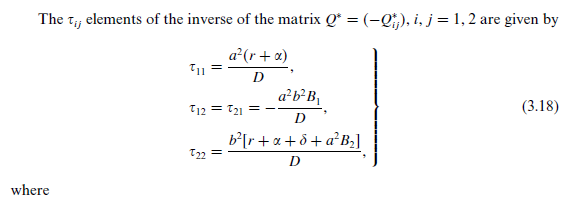
In this article, we consider Lindley’s approximation form expanding about the posterior mode. For the two parameter case *k = (k1,* k2), Lindley’s approximation leads to 

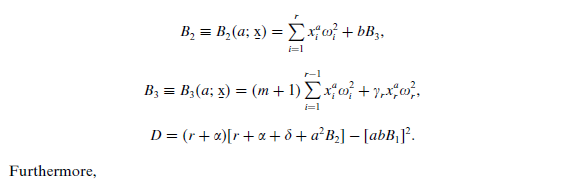
The joint posterior mode, denoted by *(a,b*), is obtained from (3.15) and is given by

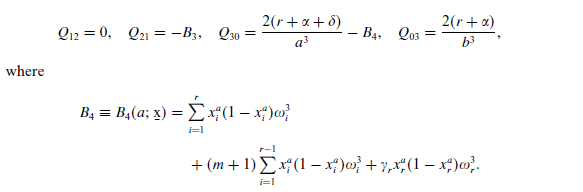


where *a* is the solution of the following nonlinear equation

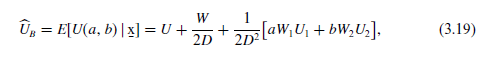


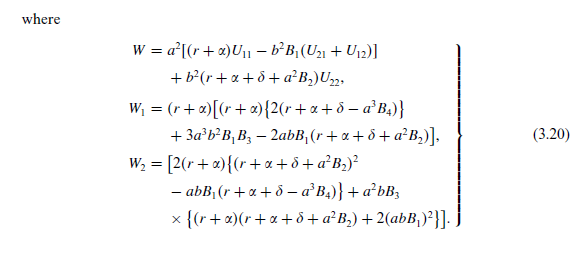






Substituting the above values in (3.14) yields the Bayes estimate of a function *U* = U(a, b), of the unknown parameters a and b, given by:





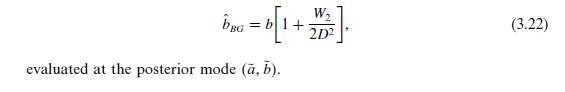
All functions in the right-hand side of (3.19) are to be evaluated at the posterior mode *(a,* b). Now, the Bayes estimates of a and *b* are computed as follows:

1. If *U(a,* b) = *a,* we have from (3.19),



evaluated at the posterior mode (**a, b**).

1. If U(a, b) = b, we have from (3.19),



Resources

<http://www.sciencedirect.com/science/article/pii/037837589400147N>

<http://www.itl.nist.gov/div898/handbook/apr/section1/apr123.htm>

<https://en.wikipedia.org/wiki/Stochastic_ordering>

<http://www.mathworks.com/help/stats/burr-type-xii-distribution.html?requestedDomain=www.mathworks.com>

<https://en.wikipedia.org/wiki/Burr_distribution>

<https://en.wikipedia.org/wiki/Failure_rate#hazard_function>

<https://books.google.com.tr/books?id=3c7IZjbGVdIC&pg=PA88&lpg=PA88&dq=ordering+of+random+k+th+record+value&source=bl&ots=nTxf_RR4Ax&sig=nlLbL7SMQlbChe081E_oKKClhWE&hl=tr&sa=X&ved=0ahUKEwiBucbN_ZjMAhVhIJoKHSC0BPAQ6AEIOTAD#v=onepage&q=ordering%20of%20random%20k%20th%20record%20value&f=false>