STATISTICAL RESEARCH REPORT
Institute of Mathematics
University of Oslo

No 1
January 1977

FIXED POINT CENCORING OF GAMMA DISTRIBUTIONS

by

Erik N. Torgersen

ABSTRACT

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed such that

$$P(X_i \leq x) = \int_0^x \lambda^p \Gamma(p)^{-1} e^{-\lambda x} dx \quad ; \quad 0 \leq x < 1$$

and

$$P(X_i = 1) = \int_1^\infty \lambda^p \Gamma(p)^{-1} e^{-\lambda x} dx$$

where $\lambda > 0$ is unknown and $p > 0$ is known. Then the statistic $S_n = \Sigma\{(1, X_i) : X_i < 1\}$ is minimal sufficient.

The case $p = 1$ was treated by the author in a 1973 research report. Generalizing some of these results K. Unni showed in a 1976 research report that this statistic is complete when p is irrational while it is not complete when p is an integer.

The purpose of this note is to show that S_n is boundedly complete if and only if it is complete and that this, in turn, holds if and only if p is irrational or p is rational $= \frac{r}{s}$ where r and s are relatively prime and $n > s$.

I. INTRODUCTION

Suppose the probability of death within the infinitesimal interval $(x,x+dx)$ is $\frac{\lambda^p}{\Gamma(p)} x^{p-1}e^{-\lambda x}dx$ where $\lambda > 0$ is unknown while $p > 0$ is a known constant. Inference on λ based on the observed lifespans of n randomly chosen individuals may be based on the sum of observations which is a complete and sufficient statistic. If, however, our experiment is obtained by only observing the times of death before a fixed time t , then the total number of deaths recorded together with the sum of lifelengths of individuals dying before time t constitutes a minimal sufficient statistic. It was shown by the author [1] that this statistic is not complete when $p = 1$ and $n \geq 2$. Generalizing this result K. Unni [3] proved that this holds for any integer $p \geq 1$. Unni proved furthermore that the statistic is complete when p is not an integer. His proof, however, is not quite correct and neither is his result. Having said that much, it must be stated that Unni's result is not far from being correct since his result, and his proof, holds whenever p is irrational.

The purpose of this note is to show that the model admits a complete and sufficient statistic if and only if the set :

$$\{e^{2ap\pi i} \quad ; \ a=0,1,2,\ldots\}$$

contains at least n points. By a simple number theoretical argument we arrive at the "complete" result on completeness for these models :

Theorem: The model consisting of n independent observations of the described type admits a complete and sufficient statistic if and only if : either p is irrational or p is rational $= \frac{r}{s}$ where the integers r and s are relatively prime and $n \leq s$.

Remark: If a model is not complete, then questions about the space of unbiased estimators of zero enter. In particular one might wonder if the model is quadratically complete [2] (this suffices for large parts of the theory of uniformly minimum variance unbiased estimators). If it is not quadratically complete, then it might still be boundedly complete (this suffices for the validity of some important results on testing theory). We shall, however, see that for the minimal sufficient models considered here, boundedly completeness is equivalent to completeness.

We shall prove this by considering convolution equations as those in [1]. A significant simplification is, however, obtained by adapting Unni's idea of considering the measure on $[0,\infty[$ with density x^{p-1} as the sum of its restrictions to $[0,1]$ and $[1,\infty]$.

The problem of finding uniformly minimum variance unbiased estimators in the incomplete models will not be treated here. This problem is treated by Unni in his paper and by the author in [1].

II. COMPLETENESS OF THE MINIMAL SUFFICIENT STATISTIC

We may without loss of generality assume that our observations are cencored at $t = 1$. (If they are cencored at $t = a$, replace λ by $a\lambda$ and apply the results proved for $\lambda = 1$.) Thus our basic observations $X_1, X_2, \ldots, X_n$ are independent and identically distributed such that :

$$P_\lambda(X_i < x) = \int_0^x \frac{\lambda^p}{\Gamma(p)} t^{p-1} e^{-\lambda t} dt \quad \text{if} \quad x < 1$$

while

$$P_\lambda(X_i = 1) = \int_1^\infty \frac{\lambda^p}{\Gamma(p)} t^{p-1} e^{-\lambda t} dt$$

$$= \frac{\lambda^p}{\Gamma(p)} A(\lambda)$$

where $A(\lambda) = \int_1^\infty t^{p-1} e^{-\lambda t} dt$.

Proceeding as in [1] we put :

$$d(x) = \begin{cases} 1 & \text{if} \quad x < 1 \\ 0 & \text{if} \quad x = 1 \end{cases}$$

$$t(x) = x d(x)$$

$$D_n = d_n(X_1, \ldots, X_n) = \sum_1^n d(X_i)$$

$$T_n = t_n(X_1, \ldots, X_n) = \sum_1^n t(X_i)$$

U = uniform distribution on $[0,1]$

δ_1 = the one point distribution in 1

U_p = the distribution on $[0,1]$ whose density w.r.t. U is $p x^{p-1}$.

Then

$$\frac{\lambda^{np}}{\Gamma(p)^n} (\Pi\, x_i^{d(x_i)})^{p-1}\, e^{-\lambda T_n}\, A(\lambda)^{n-D_n} \quad ; \quad x \in]0,1]^n$$ is a version of $dP_\lambda^n/d(U+\delta_1)^n$. It follows easily that (D_n, T_n) is minimal sufficient and that the conditional distribution of T_n given $D_n = d$ has density :

$$p^{-d}\left(1-\frac{\lambda^p}{\Gamma(p)}\, A(\lambda)\right)^{-d} \frac{\lambda^{pd}}{\Gamma(p)^d}\, e^{-\lambda t} \quad ; \; t \in [0,d] \quad \text{w.r.t.} \quad U_p^{d*}.$$

Furthermore the variable D_n is binomially distributed with success parameter $1 - \frac{\lambda^p}{\Gamma(p)} A(\lambda)$. It follows in particular that a variable $\delta(D_n, T_n)$ is integrable if and only if $\delta(d,\circ)$ is U_p^{d*} integrable for all d .

Let $\delta(D_n, T_n)$ be integrable. Then its expectation may be written :

$$E\,\delta = E\, E(\delta | D_n)$$

$$= \frac{\lambda^{np}}{\Gamma(p)^n} \sum_{d=o}^{n} \left(\frac{A(\lambda)}{A(1)}\right)^{n-d} \int_o^{\infty} e^{(1-\lambda)t} e^{-t} \delta(d,t) \binom{n}{d} p^{-d} A(1)^{n-d} U_p^{d*}(dt).$$

Introduce for any integrable variable $\delta(D_n, T_n)$ the sequence $\sigma_o, \sigma_1, \sigma_2, \dots, \sigma_n$ of measures such that σ_i is supported by $[0,i]$ and $d\sigma d/dU_p^{d*} = \binom{n}{d} A(1)^{n-d}\, e^{-t}\, \delta(d,t) p^{-d}$.

Let τ be the probability measure on $[1,\infty[$ whose density w.r.t. Lebesgue measure is

$$A(1)^{-1}\, t^{p-1}\, e^{-t} \quad ; \quad t > 0 .$$

Note first that $(\sigma_o, \sigma_1, \dots, \sigma_n)$ is, and may be, any sequence of finite measures such that σ_o is concentrated in 0 while σ_i for $i \geq 1$ is absolutely continuous and concentrated on $[0,i]$.

Furthermore the correspondence :

$$\delta <-> \sigma$$

is linear and clearly 1-1 .

Using the convolution property of Laplace transforms we see that :

$$E\,\delta = \frac{\lambda^{np}}{\Gamma(p)^n} \int_o^\infty e^{(1-\lambda)t} \Big[\sum_{d=o}^{n} \tau^{(n-d)*} * \sigma_d \Big] (dt) \ .$$

Before arguing we need one more simplification ; and this simplification is our adaption of the similar simplification in Unni's paper.

Let H and K be the probability measures on $[0,\infty[$ whose densities w.r.t. Lebesgue measure are, respectively :

$$\text{constant} \quad t^{p-1}\, e^{-t} \quad ; \quad t > 0$$

and

$$\text{constant} \quad t^{p-1}\, e^{-t}\, I_{[o,1[}(t) \quad ; \quad t > 0 \ .$$

The constants are, respectively, $\Gamma(p)^{-1}$ and $(\Gamma(p)-A(1))^{-1}$.

Put $\alpha = \Gamma(p)/A(1)$ and $\beta = (\Gamma(p)/A(1))- 1$.

Then $\tau = \alpha H - \beta K$ and α and β are positive constants depending on p only.

It follows that the convolution sum in the expression for $E\,\delta$ may be written

$$\sum_{d=o}^{n} (\alpha H-\beta K)^{(n-d)*} * \sigma_d$$

$$= \sum_{k=o}^{n} H^{(n-k)*} * R_k$$

where :

$$R_k = \sum_{d=o}^{k} \alpha^{n-k} \binom{n-d}{n-k} (-1)^{k-d} \beta^{k-d} K^{(k-d)*} * \sigma_d \ .$$

Note that R_k is supported by $[0,k]$.

By the uniqueness theorem for Laplace transforms:

$$E_\lambda \delta \equiv 0 \quad \text{if and only if :}$$

$$\sum_{d=o}^{n} \tau^{(n-d)*} * \sigma_d = 0 \quad \text{or, equivalently, that} \quad \sum_{k=o}^{n} H^{(n-k)*} * R_k = 0 \;.$$

Suppose this condition holds.

By the expression for $E\,\delta$ on page 4 we see, by dividing by $\frac{\lambda^{np}}{\Gamma(p)^n}\left(\frac{A(\lambda)}{A(1)}\right)^n$ and letting $\lambda \downarrow 0$, that $\delta(0,0) = 0$. Note also, since $\sum_{d=o}^{n-1} \tau^{(n-d)*} * \sigma_d$ is supported by $[1,\infty]$ that $\delta(n,t) = 0$ for U almost all t . [It follows in particular that, except for integrability conditions, there are no restrictions on the restrictions of UMVU estimators

to $\{(0,0)\} \cup \{(n,t) : 0 \le t \le 1\}$.]

We shall in the following restrict attention to estimators δ such that $\delta(0,0) = 0$. By the above result ; no unbiased estimator of zero is excluded.

Note next that, by the 1-1 correspondence :

$$R_k = \sum_{d=1}^{k-1} \alpha^{n-k} \binom{n-d}{n-k} (-1)^{k-d} \beta^{k-d} K^{(k-d)} * \sigma_d + \alpha^{n-k} \sigma_k \quad ; \quad k=1,2,\dots,n$$

that $(R_1, R_2, \dots, R_n)$ is and may be any sequence of absolutely continuous and finite measures such that R_i ; $i \ge 1$ is supported by $[0,i]$. Furthermore the correspondences

$$\delta \text{ <-> } \sigma \text{ <-> } R$$

are all linear 1-1 and onto.

It follows in particular that the model is complete if and only if

$$(\S) \quad \sum_{k=1}^{n} H^{(n-k)*} R_k = 0 \Rightarrow R_1 = \ldots = R_n = 0 .$$

Let us write H_p instead of H. Then $H_p^{m*} = H_{mp}$ so that (§) may be written

$$\sum_{k=1}^{n} H_{p(n-k)} * R_k = 0 \Rightarrow R_1 = \ldots = R_n = 0 .$$

Consider first the case:

p = integer

$n = 2$.

Then (§) may be written

$$H_p * R_1 + R_2 = 0 \Rightarrow R_1 = R_2 = 0$$

or equivalently:

$$H_p * R_1 \text{ is supported by } [0,2] \Rightarrow R_1 = 0 .$$

The density of $H_p * R_1$ in $y > 1$ may be written :

$$\int_0^1 H_p'(y-x)\, R_1'(x)\, dx$$

or :

$$\int_0^1 \frac{\lambda^p}{\Gamma(p)} (y-x)^{p-1}\, e^{-(y-x)}\, R_1(dx) .$$

Hence $H_p * R_1$ is supported by $[0,2]$ if and only if it is supported by $[0,1]$ and this, in turn, is equivalent to :

$$\int_0^1 (y-x)^{p-1}\, e^x\, R_1(dx) = 0 \quad ; \quad y > 1$$

or:

$$\sum_{j=0}^{p-1} \binom{p-1}{j} y^j \int_0^1 (-x)^{p-1-j}\, e^x\, R_1(dx) = 0 \quad , \; y > 1$$

or :

$$\int_0^1 x^j e^x R_1(dx) = 0 \quad , \quad j=0,1,\dots,p-1 .$$

A non vanishing measure R_1 with this property may be obtained by letting $e^x R_1$ be the difference between two different and absolutely continuous probability measures on $[0,1]$ such that the j-th moments are the same for $j \leq p-1$.

It follows that the model is incomplete for p integer and $n=2$. By Torgersen [1] the model is not complete whenever p is an integer and $n \geq 2$. This was proved by Unni in [3] . Consider next the case of a rational number $p = \frac{r}{s}$ and $n = s+1$. (§) may then be written :

$$H_r * R_1 + \dots + R_n = 0 \Rightarrow R_1 = \dots = R_n = 0 .$$

By the above result, however, we may just let $R_i = 0$, $i=2,\dots,n-1$ and let R_1 and R_n be non zero measures on $[0,1]$ such that

$$H_r * R_1 + R_n = 0 .$$

It follows by this result and the last section in [2] that the model is incomplete when $p = \frac{r}{s}$ where (r,s) are relatively prime and $n \geq s+1$.

Consider next conditions assuring completeness. The equation :

$$\sum_{k=1}^{n} H_{p(n-k)} * R_k = 0$$

may be written :

$$\int e^{(1-\lambda)t} (\Sigma)\ (dt) = 0$$

or

$$\sum_{k=1}^{n} \lambda^{pk} \hat{R}_k(\lambda) \quad ; \lambda > 0$$

where $\hat{R}_k(\lambda) = \int e^{(1-\lambda)t} R_k(dt)$ is, for each k, an entire function. Put $\zeta = e^{2p\pi i}$. Inserting $\lambda = e^z$ we find

$$\sum_{k=1}^{n} e^{pkz} \hat{R}_k(e^z) \equiv 0 .$$

Replacing z by $z+a\circ 2\pi i$ where a is an integer, we get :

$$\sum_{k=1}^{n} \zeta^{ak} e^{pkz} \hat{R}_k(e^z) \equiv 0 .$$

Now $\zeta^a = \zeta^b$ if and only if $p(a-b)$ is an integer. Furthermore :

$$\begin{vmatrix} z_1 & \dots & z_n \\ z_1^2 & \dots & z_n^2 \\ z_1^n & \dots & z_n^n \end{vmatrix} = \Pi(z_j - z_i) \quad \text{where} \quad z_o = 0$$

and the product is taken over all pairs (i,j) such that $0 \leq i < j \leq n$. Thus this determinant differs from zero if and only if the numbers $z_1, \dots, z_n$ are all distinct and different from zero. In particular

$$\begin{pmatrix} \zeta^{a_1} & \cdot & \zeta^{na_1} \\ \vdots & \vdots & \\ \zeta^{a_n} & \cdot & \zeta^{na_n} \end{pmatrix} \quad \text{is non singular whenever}$$

$$\zeta^{a_1}, \zeta^{a_2}, \dots, \zeta^{a_n} \quad \text{are all different.}$$

If p is irrational then we may take $a_i = i$; $i=0,1,2,\dots,n-1$. If $p = \frac{r}{s}$ where $(r,s) = 1$ then $\zeta^a = \zeta^b$ if and only if $a \equiv b \mod (s)$ and we may take $a_i = i$, $i=0,1,\dots,n-1$ whenever $n \leq s$.

In any case we find that :

$$e^{pkz} \hat{R}_k(e^z) \equiv 0 \quad ; \; k=1,\dots,n$$

i.e. : $R_k = 0$, $k=1,2,\ldots,n$.

It follows that the model is complete in these cases. Altogether we have proved :

<u>Theorem.</u> The class of possible distributions for (D_n,T_n) is not complete if and only if p is rational $= \frac{r}{s}$ where r and s are relatively prime and $n > s$.

Let us finally consider the problem of extremality of these models. According to [2] a model is extremal if and only if it admits a sufficient and boundedly complete statistic. Thus the following is a strengthening of the theorem above :

<u>Theorem.</u> Let $X_1,X_2,\ldots,X_n$ be independent and identically distributed such that :

$$P(X_i \leq x) = \int_0^x \lambda^p \, \Gamma(p)^{-1} \, x^{-\lambda x} \, dx \quad ; \quad 0 \leq x < 1$$

and

$$P(X_i = 1) = \int_1^\infty \lambda^p \, \Gamma(p)^{-1} \, x^{p-1} \, e^{-\lambda x} \, dx \, .$$

Suppose $\lambda > 0$ is unknown and that $p > 0$ is known. Then the experiment $\mathcal{E}_p^n$ defined by $X_1,\ldots,X_n$ admits a complete and sufficient statistic if and only if it is extremal and this is the case if and only if either

i) p is irrational

or

ii) p is rational $= \frac{r}{s}$ where the integers r and s are relatively prime and $n \leq s$.

The proof is based on the following proposition :

<u>Proposition:</u> Let δ be any everywhere integrable function of (D_n,T_n). Then the following conditions are all equivalent :

(i) δ is essentially bounded.

(ii) For some constant $M < \infty$: $|\sigma_k| \leq M\, U_p^{k*}$; $k=0,1,\ldots,n$.

(iii) For some constant $M < \infty$: $|R_k| \leq M\, U_p^{k*}$; $k=0,1,\ldots,n$.

Proof: (i) and (ii) are equivalent by the definition of σ . Suppose (ii) holds. Then, by the definition of R :

$$|R_k| \leq \sum_{d=o}^{k} \text{constant} \cdot K^{(k-d)*} * |\sigma_d|$$

$$\leq \text{constant} \cdot \sum_{d=o}^{k} K^{(k-d)*} * U_p^{d*}$$

$$\leq \text{constant} \cdot (K+U_p)^{k*} .$$

Now $[dK/dU_p]_t = \text{constant} \cdot e^{-t} \leq \text{constant}.$

Hence $|R_k| \leq \text{constant} \cdot U_p^{k*}$ so that (iii) holds. Suppose finally that (iii) is satisfied. As

$d\sigma_o/d\,U_p^{o*} = A(1)^n\, \delta(0,0)$ and $|\delta(0,0)| < \infty$

we have always $|\sigma_o| \leq \text{constant} \cdot U_p^{o*}$. By

$R_1 = \alpha^{n-1}[(-n)\, \beta K*\sigma_o + \sigma_1]$ we get :

$\sigma_1 = \text{constant} \cdot R_1 + \text{constant} \cdot K$ so that :

$|\sigma_1| \leq \text{constant} \cdot U_p + \text{constant} \cdot U_p \leq \text{constant} \cdot U_p$.

We proceed by induction. Suppose

$|\sigma_{k'}| \leq M\, U_p^{k'*}$ when $k' < k$. By the definition of R_k :

$$\sigma_k = \sum_{d=o}^{k-1} \text{constant} \cdot K^{(k-d)*} * \sigma_d$$

$$+ \ (\text{constant}) \cdot R_k$$

so that :

$$|\sigma_k| \leq \sum_{d=o}^{k-1} \text{constant} \cdot U_p^{(k-d)*} * U_p^{d*}$$

$$+ \text{constant} \cdot U_p^{k*} \leq \text{constant} \cdot U_p^{k*} .$$

□

Proof of the theorem: We must show that whenever the model is not complete it is actually boundedly incomplete. Consider then the proof above for the case p = integer and $n \geq 2$. The problem was reduced to find an absolutely continuous and non-vanishing measure R_1 on $[0,1]$ such that:

$$\int_o^1 x^j e^x R_1(dx) = 0 \quad ; \quad j=0,1,\ldots,p-1 .$$

Let $q > 0$. We shall now argue that R_1 may be chosen so that $R_1 \leq \text{constant } U_q$. Assume namely that this is not the case and let $0 < a < b < 1$. Let F_j be the probability distribution on $[a,b]$ with density=constant x^j. Then, by the boundedness of e^x on $[0,1]$, the experiment $(F_o,F_1,\ldots,F_{p-1})$ is boundedly complete. This, in turn, implies by an argument of Le Cam [see 2] that the class of polynomials on $[a,b]$ of degree $\leq p-1$ is fundamental in $L_1(a,b)$. We have thus, as this is not the case, arrived at a contradiction.

It follows that we may choose R_1 so that $|R_1| \leq \text{constant } U_p$ and then, since $R_2 = - H_p * R_1$: $|R_2| \leq \text{constant} \cdot H_p * U_p$ so that on $[0,2]$:

$$|R_2| \leq \text{constant} \cdot U_p * U_p = \text{constant} \cdot U_p^{2*} .$$

This imply, by the proposition above, that the model is not boundedly complete when p is an integer and $n \geq 2$. Consider finally the case where p is rational , $p = \frac{r}{s}$ where r and s

are integers such that $(r,s) = 1$ and $n = s+1$. By the result described above we may choose non vanishing, absolutely continuous and finite measures R_1 and R_2 such that both are supported by $[a,1]$ and such that

$$H_r * R_1 + R_2 = 0 \quad \text{and} \quad |R_1| \leq \text{constant } U_p .$$

It remains, by the proof of the first theorem, to show that $|R_2| \leq$ constant U_p^{n*} on $[a,1]$. This, however, follows since the density of $R_2 = H_r * (-R_1)$ may be chosen continuous on $[a,1]$ while the density of U_p^{n*} may, since $n \geq 2$, be chosen positive and continuous on $[a,1]$.

References.

[1] Torgersen, E.N. (1973). Uniformly minimum variance unbiased (UMVU) estimators based on samples from right truncated and right accumulated exponential distributions. Statistical Research Report. Univ. of Oslo.

[2] Torgersen, E.N. (1977). Mixtures and products of dominated experiments. Ann.Statist. January 1977.

[3] Unni, K. (1976). Unbiased estimation in type I cencored gamma distributions. Research Report. Indian Statistical Institute, Calcutta.