On Statistics Independent of a Sufficient Statistic: Basu's Lemma

UWE KOEHN AND D. L. THOMAS*

1. *Introduction*

In 1955 D. Basu [1] published a proof of the statement that if $\mathbf{P} = \{P_\theta : \theta \in \Omega\}$ is a family of probability measures on an abstract sample space $(\mathbf{X}, \mathbf{A})$ and S is a sufficient statistic for $\mathbf{P}$, then for a statistic V to be stochastically independent of S for every θ the probability distribution of V must be free of θ. Since that time his theorem has been given in many courses on advanced statistics and has appeared in several textbooks, e.g., Vol. 2 of *The Advanced Theory of Statistics* by M. G. Kendall and A. Stuart [3]. Although the result has not had the prominent position in statistics which its partial converse (i.e., the fact that when S is a boundedly complete sufficient statistic the distribution of V being free of θ implies V is independent of S) has had, it has been used to prove characterizations of certain families of distributions, e.g., Basu [1].

Subsequently Basu, Farrell and others recognized that Basu's original statement needed qualification and Basu [2] gave a sufficient condition for its validity. Nevertheless, since many statisticians do not seem to have realized the original error and because we have a simple proof of necessary and sufficient conditions for the validity of Basu's result, we have submitted this note.

In Section 2 we give three simple examples. The simplest, given in Example 2, suffices to show the need for additional conditions; however, since some readers may understand the problem more clearly in the context of continuous distributions, in Example 1 we give a counterexample involving uniform distributions. The third example illustrates when Basu's statements will hold. In the final section we give a corrected version of Basu's theorem.

2. *Examples*

Example 1: Suppose we take $\mathbf{X}$ to be the real line, $\mathbf{A}$ to be the Borel sets, and P_θ to be uniform measure over the half-open interval $[\theta, \theta + 1)$ where θ is an integer. Since we can write the density functions in the form

$$P_\theta(x) = I_{(\theta,\theta+1)}(x) = \begin{Bmatrix} 1 & \text{if} & [x] = \theta \\ 0 & \text{if} & [x] \neq \theta \end{Bmatrix} = I_{(\theta,\theta)}([x])$$

with $[x]$ defined to be the largest integer less than or equal to x, we know by the factorization theorem that $S = [X]$ is a sufficient statistic for the family $\mathbf{P}$. For each A in $\mathbf{A}$ and any integers, θ and $\mathbf{s}$,

$$P_\theta(X \in A, S(X) = \mathbf{s}) = \begin{cases} P_\theta(X \in A) & \text{if} \quad \mathbf{s} = \theta \\ 0 & \text{if} \quad \mathbf{s} \neq \theta \end{cases}$$
$$= P_\theta(X \in A)\, P_\theta(S(X) = \mathbf{s});$$

hence, X and S are independent for every θ. Of course, the distribution of X depends on θ and Basu's original claim fails to hold here. We might note in passing that S is a constant with probability one for each θ so that it is actually independent of any other statistic, and it is complete.

Example 2: To obtain another counterexample let $\mathbf{P} = \{P_1, P_2\}$ be a family of two probability measures on $(\mathbf{X}, \mathbf{A})$ such that $P_2(A) = P_1(A^c) = 1$ and $P_2(A^c) = P_1(A) = 0$ for some A in $\mathbf{A}$. Take $S(\cdot) = I_A(\cdot)$ to be the indicator function of the set A and write $\mathbf{P}^S = \{P_1{}^S, P_2{}^S\}$ for the family of induced probability measures on the image space under S. It is clear that

$$P_2(X \in A \mid S(X) = 1) = P_1(X \in A^c \mid S(X) = 0) = 1$$

and

$$P_1(X \in A \mid S(X) = 0) = P_2(X \in A^c \mid S(X) = 1) = 0;$$

hence, for each θ in $\Omega = \{1, 2\}$ we have

$$P_\theta(X \in A \mid S(X) = \mathbf{s}) = \mathbf{s} \qquad \text{a.e.} \quad (P_\theta{}^S)$$

and

$$P_\theta(X \in A^c \mid S(X) = \mathbf{s}) = 1 - \mathbf{s} \qquad \text{a.e.} \quad (P_\theta{}^S).$$

Therefore, $S = S(X)$ is a sufficient statistic for the family $\mathbf{P}$. Now for any A' in $\mathbf{A}$ and each θ in Ω

$$\begin{aligned} P_\theta(X \in A', S(X) = 1) &= P_\theta(X \in A', X \in A) \\ &= \begin{cases} P_2(X \in A' \cap A) & \text{if} \quad \theta = 2 \\ 0 & \text{if} \quad \theta = 1 \end{cases} \\ &= \begin{cases} P_2(X \in A') & \text{if} \quad \theta = 2 \\ 0 & \text{if} \quad \theta = 1 \end{cases} \\ &= P_\theta(X \in A') P_\theta(S(X) = 1) \end{aligned}$$

and, similarly,

$$P_\theta(X \in A', S(X) = 0) = P_\theta(X \in A') P_\theta(S(X) = 0).$$

Thus, X and S are independent for each θ even though the distribution of X depends on θ. So, Basu's theorem is not true without restrictions on the family $\mathbf{P}$ under consideration.

Example 3. It will follow from the theorem in the next section that we could enlarge the parameter space of

* Dept. of Statistics, Univ. of Connecticut, Storrs, Conn. 06268.

Example 1 somewhat and still be able to find a statistic independent of the sufficient statistic whose distribution depends on θ; however, if enough of the "right" parameter points are added to the original space, then Basu's theorem will be correct. For example, allow θ to range over the set $\{k/2: k \text{ is an integer}\}$. Then, via the factorization theorem the statistic defined by

$$S(x) = k \qquad \text{if} \qquad \frac{k}{2} \le x < \frac{k+1}{2}$$

can be shown to be sufficient. But, now,

$$P_\theta(X \in A, S(X) = \mathbf{s}) = \begin{cases} P_\theta(X \in A \cap [\theta, \theta + \tfrac{1}{2})) & \text{if } \mathbf{s} = 2\theta \\ P_\theta(X \in A \cap [\theta + \tfrac{1}{2}, \theta + 1)) & \text{if } \mathbf{s} = 2\theta + 1 \\ 0 & \text{elsewhere,} \end{cases}$$

which is not always equal to

$$P_\theta(X \in A)P_\theta(S(X) = \mathbf{s}) = \begin{cases} P_\theta(X \in A)/2 & \text{if } \mathbf{s} = 2\theta \text{ or } \mathbf{s} = 2\theta + 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Thus, as an illustration of the fact that for this family $\mathbf{P}$ any statistic whose distribution depends on θ will not be independent of S for all θ, we have X and S dependent. The important change is the destruction of the disjointness of the distributions in Example 1.

3. *Sufficiency and Independence*

To understand the difficulty with the original "proof" of Basu's result let $\{P_\theta\}$ be a family of probability measures for which S is a sufficient statistic and suppose V is a statistic that is independent of S for all θ. If $(\mathbf{S}, \mathbf{B}, P_\theta{}^S)$ is the range space of S, then

$$\begin{aligned} P_\theta(V \in C, S \in B) &= \int_B P(V \in C|\mathbf{s})\, dP_\theta{}^S \\ &= P_\theta(V \in C)P_\theta(S \in B) \\ &= P_\theta(V \in C) \int_B dP_\theta{}^S, \end{aligned}$$

where we drop the θ on $P(V \in C \mid \mathbf{s})$ because S is sufficient. Since

$$\int_B P(V \in C \mid \mathbf{s})P_\theta{}^S = \int_B P_\theta(V \in C)\, dP_\theta{}^S$$

for every B in $\mathbf{B}$, then

$$P(V \in C \mid \mathbf{s}) = P_\theta(V \in C) \quad \text{a.e.} \quad (P_\theta{}^S)$$

for each θ, i.e., except for $\mathbf{s} \in B_\theta$ where $P_\theta{}^S(B_\theta) = 0$ we have $P(V \in C \mid \mathbf{s}) = P_\theta(V \in C)$. The mistake usually made consists of concluding that, therefore,

$$P_{\theta_1}(V \in C) = P(V \in C \mid \mathbf{s}) = P_{\theta_2}(V \in C)$$

which may not be true because B_θ may vary with θ. In Example 2, for instance, B_1 is the complement of B_2.

The following theorem and corollary give necessary and sufficient conditions for the result under discussion to hold.

Theorem: Let $\{P_\theta\}$ be a family of probability measures on the measurable space $(\mathbf{X}, \mathbf{A})$ for which S is a sufficient statistic. There exists a statistic V, independent of S for all θ, whose distribution depends on θ if, and only if, there exists a set $A \in \mathbf{A}$ such that $P_\theta(A) = 0$ or 1 for all θ with $P_\theta(A) = 1$ for some θ and $P_\theta(A) = 0$ for some θ. In that case we will call A a splitting set.

Proof: We recall that if a random variable is a constant with probability one, then it is independent of all random variables.

Now suppose A is a splitting set and let $\omega = \{\theta: P_\theta(A) = 1\}$. If $V(x) = I_A(x)$, the indicator function of the set A, then we see that $P_\theta(V(x) = 1) = 1$ for $\theta \in \omega$ and $P_\theta(V(x) = 0) = 1$ for $\theta \notin \omega$. Thus, the distribution of V depends on θ but V is a constant with probability one for each θ and so it is independent of S.

On the other hand, let V be a random variable independent of S for all θ whose distribution depends on θ so that there exists a measurable set C such that

$$P_{\theta_1}(V \in C) \neq P_{\theta_2}(V \in C) \tag{1}$$

for some parameter values θ_1 and θ_2. Since S is sufficient

$$P(V \in C \mid S(x)) = P_\theta(V \in C) \quad \text{a.e.} \quad (P_\theta, \mathbf{A}_S) \tag{2}$$

where $\mathbf{A}_S$ is the sub-σ-field induced by S. To emphasize that $P(V \in C \mid S(x))$ is a function of x, let $h(x) = P(V \in C \mid S(x))$. Also let

$$A_\theta = \{x: h(x) = P(V \in C \mid S(x)) = P_\theta(V \in C)\}.$$

Now A_θ is the set on which $h(x)$ takes the value $P_\theta(V \in C)$ and (1) says that there are at least two such values. Since h is a function, the sets A_θ corresponding to different values are distinct and those corresponding to identical values are equal. So $A_{\theta_1} \cap A_{\theta_2} = \emptyset$. By (2), $P_\theta(A_\theta) = 1$ for every θ.

Let $A = A_{\theta_1}$ and $\omega = \{\theta: A_\theta = A\}$. Of course, $\theta_1 \in \omega$ and $\theta_2 \in \Omega - \omega$ and so neither ω nor $\Omega - \omega$ is empty. If $\theta \in \omega$, then $P_\theta(A) = P_\theta(A_\theta) = 1$ and if $\theta \in \Omega - \omega$, $P_\theta(A^c) \geq P_\theta(A_\theta) = 1$, since $A_\theta \subset A^c$ (i.e., if $x \in A_\theta$ then $h(x) = P_\theta(V \in C) \neq P_{\theta_1}(V \in C)$ and so $x \notin A_{\theta_1} = A$). Hence, A is a splitting set.

Corollary: Let $\{P_\theta\}$ be a family of probability measures on $(\mathbf{X}, \mathbf{A})$ and let S be sufficient for the family. Every statistic independent of S for all θ has a distribution that does not depend on θ if, and only if, there does not exist a measurable splitting set A, i.e., a set such that $P_\theta(A) = 1$ for various θ and $P_\theta(A) = 0$ for all other θ, neither set of θ being vacuous.

Proof: This statement is merely the contrapositive of the theorem.

Remark 1: Basu's [1] proof of his characterization of the normal distribution is not valid as it stands. Extra conditions such as the nonexistence of a splitting set A are needed.

Remark 2: Basu [2] gives a sufficient condition for any statistic V independent of S to have a distribution free of the parameter θ. His condition is implied by ours but the reverse implication does not hold.

Basu defines the overlapping of two probability measures, μ_a and μ_b, on $(\mathbf{X}, \mathbf{A})$ by the condition that if $B \in \mathbf{A}$ and $\mu_a(B) = 1$, then $\mu_b(B) > 0$. Two measures, μ_a and μ_b, in a family are said to be connected if there exists a finite sequence of probability measures in the family, $\mu_0 = \mu_a, \mu_1, \ldots, \mu_k = \mu_b$, such that μ_i and $\mu_{i+1} (i = 0, 1, \ldots, n - 1)$ overlap. It is clear that if every two probability measures in a family are connected, then no splitting set A exists that divides the family into two classes, and so having a connected family is a sufficient condition.

On the other hand consider the Borel sets on the real line and let N be a non-measurable set. Let $\mathbf{P}$ be the family of two point measures putting probability $\frac{1}{2}$ on each point with the restriction that both points are in N or in its complement. Clearly connectedness is impossible since no measure concentrated in N overlaps with any concentrated in the complement of N. However the only splitting set is N which is not measurable. Thus any statistic independent of a complete sufficient statistic has a distribution free of the parameter while Basu's condition does not hold. Actually in this situation a single observation, S, is complete and sufficient since N and its complement each contain at least three points.

Acknowledgment: The example in Remark 2 was suggested to the authors by Professor Stuart Sidney of the University of Connecticut Mathematics Department. The authors also wish to thank the referees for helpful and stimulating comments.

REFERENCES

[1] Basu, D., "On Statistics Independent of a Complete Sufficient Statistic," *Sankhyā*, Ser. A, 15 (1955), 377–380.

[2] ———, "On Statistics Independent of Sufficient Statistics," *Sankhyā*, Ser. A, 20 (1958), 223–226.

[3] Kendall, M. G. and Stuart, A., *The Advanced Theory of Statistics*, Vol. 2, New York: Hafner Publishing Company, 1967.

From the Noncentral *t* to the Normal Integral

DOUGLAS M. HAWKINS*

The noncentral t distribution with n degrees of freedom and noncentrality δ is of great importance in hypothesis testing as it yields the power of the widely used t tests (see for example Hogg and Craig [2, p 292], Kendall and Stuart [3, p 264]). Pedagogically, too, it is quite easily derived by standard transformation methods or, rather less usually, as the compound distribution resulting when a $N(\delta v^{-1}, v^{-2})$ distribution is mixed by letting nv^2 follow a $\chi_n{}^2$ distribution.

Its complementary cumulative distribution function is

$$P[T > t \mid n, \delta] = \frac{\exp(-\frac{1}{2}\delta^2)}{(\pi n)^{1/2}\Gamma(\frac{1}{2}n)} \sum_{j=0}^{\infty} \frac{\Gamma\{\frac{1}{2}(n+j+1)\}}{j!} \times (\delta\sqrt{2})^j \int_t^\infty \frac{x^j}{n^{1/2j}[1 + x^2/n]^{1/2(n+j+1)}}\, dx \qquad (1)$$

This expression may be simplified in several stages. First, the integral is written in the standard form for an incomplete beta function by making the change of variable

$$u = n/(n + x^2).$$

Next, the term in $j!$ is rewritten using the duplication formula for gamma functions

$$j! = \Gamma(j+1) = 2^j\Gamma\{\tfrac{1}{2}(j+1)\}\Gamma(\tfrac{1}{2}j+1)/\sqrt{\pi}.$$

These transformations bring (1) to the form

$$\frac{\exp(-\frac{1}{2}\delta^2)}{\Gamma(\frac{1}{2}n)\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{\Gamma\{\frac{1}{2}(n+j+1)\}\sqrt{\pi}}{2^j\Gamma\{\frac{1}{2}(j+1)\}\Gamma(\frac{1}{2}j+1)} (\delta\sqrt{2})^j.$$

$$\tfrac{1}{2}\int_0^{u(t)} u^{1/2n-1}(1-u)^{1/2(j-1)}\, du. \qquad (2)$$

Introducing the incomplete beta ratio

$$I_x(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}\, dt,$$

we see that (2) may be written in the form

$$\tfrac{1}{2}e^{-1/2\delta^2} \sum_{j=0}^{\infty} \frac{(\delta/\sqrt{2})^j}{\Gamma(\frac{1}{2}j+1)} I_\alpha\{\tfrac{1}{2}n, \tfrac{1}{2}(j+1)\}, \qquad (3)$$

where $\alpha = n/(n + t^2)$.

The recursion that results from integrating $I_x(a, b)$ by parts makes the representation (3) suitable for computer implementation. A very similar expansion in terms of incomplete beta ratios may also be found for the noncentral F distribution, whose density is given by Anderson [1] p 114.

A rather surprising result emerges from (3) if we set $t = 0$. In this case $\alpha = 1$ and $I_1(a, b) = 1$ for all a and b. Thus

* Dept. of Applied Mathematics, Univ. of the Witwatersrand, 1 Jan Smuts Ave., Johannesburg 2001, South Africa.