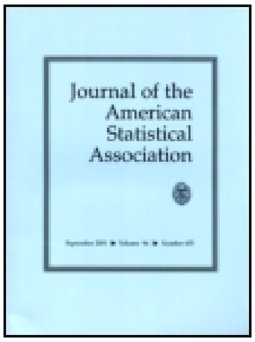

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

A Note on the Influence of the Sample on the Posterior Distribution

Ward Whitt

To cite this article: Ward Whitt (1979) A Note on the Influence of the Sample on the Posterior Distribution, Journal of the American Statistical Association, 74:366a, 424-426, DOI: 10.1080/01621459.1979.10482530

To link to this article: https://doi.org/10.1080/01621459.1979.10482530

A Note on the Influence of the Sample on the Posterior Distribution

WARD WHITT*

This article studies when and how the posterior distribution responds monotonically to changing sample evidence. The monotone likelihood ratio (MLR) property, which implies stochastic order, is suggested as a convenient ordering to express and demonstrate this monotonicity.

KEY WORDS: Bayesian inference; Posterior distribution; Monotone likelihood ratio; Stochastic order.

1. INTRODUCTION

Suppose we sample at random without replacement from an urn containing N balls, each of which is colored either red or white. Given our prior distribution of the number of red balls in the urn and the observed number of red balls in the sample, we can compute using Bayes's theorem the posterior distribution of the number of red balls either in the entire population or remaining in the urn after the sample has been removed. We usually expect the sample to reflect the population in the sense that the more we see in the sample, the more we believe we will see in the population. When we speak of the posterior distribution on the entire population, this statement is always true in a strong sense; see Theorems 2 and 4. We also usually expect this statement to be true when we speak of the posterior distribution on the unsampled population. This statement, however, obviously need not always be true. For example, if we know in advance that exactly k of the N balls are red, then the number of red balls remaining in the urn decreases as the number of red balls in the sample increases. We provide conditions on the prior distribution for the posterior distribution on the unsampled population to be an increasing function (in some sense) of the sample evidence; see Theorem 3. The monotonicity properties discussed here are similar to convergence properties of the posterior distribution as the sample size increases (DeGroot 1970, p. 201, and references there). We do not know of any earlier work on the kind of monotonicity treated here, but such properties are not difficult to deduce in the case of conjugate prior distributions on real parameters (DeGroot 1970, Ch. 9). For example, suppose $X_1, \ldots, X_n$ is a random sample from a Bernoulli distribution with an unknown value of the parameter W. Suppose also that the prior distribution of W is a beta distribution with parameters α and β such that $\alpha > 0$ and $\beta > 0$. Then the posterior distribution of W when $X_k = x_k (k = 1, \ldots, n)$ is a beta distribution with parameters $\alpha + y$ and $\beta + n - y$, where $y = x_1 + \ldots + x_n$ (DeGroot 1970, p. 160). Hence, the mean of the posterior distribution is $(\alpha + y)/(\alpha + \beta + n)$, which is clearly strictly increasing in y for fixed n and strictly decreasing in n for fixed y.

2. COMPARING PROBABILITY DISTRIBUTIONS

The sense in which the posterior distribution will be shown to be an increasing function of the sample is important. We wish to show not only that the mean increases but that the whole distribution increases. For this purpose, we use the following two partial-order relations on probability mass functions. Suppose $p \equiv \{p(k); k = 0, 1, \ldots, N\}$ and $q \equiv \{q(k); k = 0, 1, \ldots, N\}$ are two probability mass functions on the set $S = \{0, 1, \ldots, N\}$.

Definition 1: The mass function p is stochastically less than or equal to the mass function q, denoted by $p \le_{st} q$, if

$$\sum_{k=m}^{N} p(k) \le \sum_{k=m}^{N} q(k) \quad \text{for all } m \ , \quad 0 \le m \le N \ .$$

Definition 2: The mass function p is less than or equal to the mass function q in the sense of monotone likelihood ratio (MLR), denoted by $p \le_r q$, if $q(k)/p(k)$ is a nondecreasing function of k (excluding k such that $p(k) = q(k) = 0$).

Stochastic order is important in decision analysis because $p \le_{st} q$ means that every decision maker who prefers more to less (has a nondecreasing utility function) prefers q to p. For further discussion about stochastic order, see Kamae, Krengel, and O'Brien (1977) or Veinott (1965, p. 769); for further discussion about the MLR property, see Ferguson (1967, p. 208) and references there. We find the relation $\le_r$ more convenient to work with than $\le_{st}$, and it is even stronger, that is, if $p \le_r q$, then $p \le_{st} q$. In fact, the two relations can be connected in an interesting way. For this purpose, let $p(A) = \sum_{k \in A} p(k)$ for any subset A of S and let p_A be the conditional probability mass function given A, defined as usual by $p_A(k) = p(k)/p(A)$, $k \in A$, for $p(A) > 0$.

* Ward Whitt is Member of Technical Staff, Bell Laboratories, Holmdel, NJ 07733. The author wishes to thank J.S. Kaufman for raising this monotonicity issue and R.E. Thomas and the referee for helpful comments and corrections.

June 1979, Volume 74, Number 366
Theory and Methods Section

Theorem 1: The MLR-ordering $p \leq_r q$ holds if and only if $p_A \leq_{st} q_A$ for all subsets A for which $p(A) > 0$ and $q(A) > 0$.

Proof: (if) Pick any k_1 and k_2 such that $k_1 < k_2$ and $p(k_i) = q(k_i) = 0$ does not occur for $i = 1, 2$. One may easily see that $p_A \leq_{st} q_A$ for $A = \{k_1, k_2\}$ implies $q(k_1)/p(k_1) \leq q(k_2)/p(k_2)$.

(only if) Note that $q_A(k)/p_A(k)$ is nondecreasing in k for all A if $q(k)/p(k)$ is nondecreasing in k. Let k_0 be such that

$$q_A(k_0)/p_A(k_0) \geq 1 > q_A(k)/p_A(k) \quad , \quad k < k_0 \ .$$

To see that k_0 exists, suppose $q_A(k)/p_A(k) < 1$ for all k. Then

$$\sum_{k\in A} q_A(k) = \sum_{k\in A} [q_A(k)/p_A(k)]p_A(k) < \sum_{k\in A} p_A(k) = 1 \quad ,$$

which contradicts q_A's being a probability mass function on A. Now, for $k_1 \geq k_0$,

$$\sum_{k\geq k_1} q_A(k) = \sum_{k\geq k_1} [q_A(k)/p_A(k)]p_A(k) \geq \sum_{k\geq k_1} p_A(k)$$

and, for $k_1 < k_0$,

$$\sum_{k\leq k_1} q_A(k) = \sum_{k\leq k_1} [q_A(k)/p_A(k)]p_A(k) < \sum_{k\leq k_1} p_A(k) \ .$$

Hence, $p_A \leq_{st} q_A$ as claimed.

Corollary: If $p \leq_r q$, then $p \leq_{st} q$.

We call the property in Theorem 1 *uniform conditional stochastic order*. We study it further in Whitt (1980).

3. THE INFLUENCE OF SAMPLING

In order to formalize the sampling, let X_i equal 1 if the ith sampled ball is red and 0 otherwise. Because we sample at random, we have N exchangeable random variables $X_1, \ldots, X_N$, each of which can assume the values 0 and 1. Let $S_n = X_1 + \ldots + X_n$, $n = 1, \ldots, N$, with $S_0 = 0$. Let the prior distribution be $P_N(k) = P(S_N = k)$, $k = 0, 1, \ldots, N$. In the following theorem, we use the MLR-ordering $\leq_r$ to describe the influence of the sample on the posterior distribution of the entire population.

Theorem 2: The conditional probability distribution on the entire population $P(S_N = \cdot | S_n = j)$ is always MLR nondecreasing in j and MLR nonincreasing in n.

Proof: By basic properties of conditional probabilities,

$$P(S_N = k | S_n = j) = \frac{f(k|N, n, j)}{\sum_{m=j}^{m=N-n+j} f(m|N, n, j)} \ , \quad k \geq j \ ,$$

where

$$f(m|N, n, j) = p_N(m) \binom{m}{j}\binom{N-m}{n-j} \bigg/ \binom{N}{n}.$$

Because the denominator is independent of k, it suffices to consider only the ratio of the numerators. In particular,

$$\frac{f(k|N, n, j+1)}{f(k|N, n, j)} = \frac{(k-j)(n-j)}{(j+1)(N-k-n+j+1)} \ ,$$

which is increasing in k. Similarly,

$$\frac{f(k|N, n+1, j)}{f(k|N, n, j)} = \frac{(n+1)(N-k-n+j)}{(N-n)(n+1-j)} \ ,$$

which is always decreasing in k.

Now we consider the influence of the sample on the unsampled population.

Theorem 3:

(i) The posterior probability distribution on the unsampled population $P(S_N - S_n = \cdot | S_n = j)$ is MLR nondecreasing (MLR nonincreasing) in j if and only if

$$\alpha_N(k) = (k+1)p_N(k+1)/(N-k)p_N(k)$$

is nondecreasing (nonincreasing) in k;

(ii) $P(S_N - S_n = \cdot | S_n = n - j, S_N - S_n \leq m < N - n)$ is MLR nondecreasing in n if and only if

$$\beta_N(k) = (k+1)(N-k-j)p_N(k+1)/(N-k)p_N(k)$$

is nondecreasing in k;

(iii) $P(S_N - S_n = \cdot | S_n = n - j)$ is MLR nonincreasing in n if and only if $\beta_N(k)$ is nonincreasing in k.

Proof:

(i) Substituting $S_N - S_n$ for S_N in the proof of Theorem 2, we obtain

$$P(S_N - S_n = k | S_n = j) = \frac{g(k|N, n, j)}{\sum_{m=0}^{m=N-n} g(m|N, n, j)} \ ,$$

where $g(m|N, n, j) = f(j+m|N, n, j)$, and

$$\frac{g(k|N, n, j+1)}{g(k|N, n, j)} = \frac{P_N(j+k+1)(j+k+1)(n-j)}{P_N(j+k)(N-j-k)(j+1)} \ ,$$

which establishes (i).

(ii) and (iii) Finally,

$$\frac{g(k|N, n+1, n+1-j)}{g(k|N, n, n-j)} = \frac{P_N(k+n+1-j)(k+n+1-j)(N-k-n)}{P_N(k+n-j)(n+1-j)(N-k-n+j)}.$$

Remark: One may easily see that the extra condition $\{S_N - S_n \leq m < N - n\}$ in (ii) is necessary. Clearly,

$$P(S_N - S_{n+1} \geq N - n | S_{n+1} = n + 1 - j) = 0 < P(S_N - S_n \geq N - n | S_n = n - j) \ .$$

Corollary: The posterior probability distribution on the unsampled population $P(S_N - S_n = \cdot | S_n = j)$ is independent of j for all n if and only if the prior is binomial, that is, $P_N(k) = (N!/k!(N-k)!)p^k(1-p)^{N-k}$ for some p.

Proof: By Theorem 3(i), $P(S_N - S_n = \cdot | S_n = j)$ is simultaneously MLR nondecreasing and MLR nonincreasing in j (and thus independent of j) if and only if $\alpha_N(k)$ is constant, which occurs if and only if

$$\begin{aligned} p_N(k) &= c(N-k+1)p_N(k-1)/k \ , \quad 1 \leq k \leq N \\ &= c^k N! p_N(0)/(N-k)!k! \end{aligned}$$

for some $c > 0$. Without loss of generality, let $c = p/(1 - p)$ for $0 \leq p \leq 1$. Then

$$p_N(k) = \binom{N}{k} p^k (1 - p)^{-k} p_N(0) \ .$$

In order to have $p_N(0) + p_N(1) + \ldots + p_N(N) = 1$, we must have $p_N(0) = (1 - p)^N$, which makes $p_N(\cdot)$ binomial with parameters N and p.

Remark: The corollary concludes that there are prior distributions for which sampling provides absolutely no additional information about the unsampled population. The binomial priors correspond precisely to the assumption of independence, which of course means the sample should have no influence on the probability distribution of the unsampled population. This phenomenon has obvious implications for situations such as destructive testing in which interest is focused on the unsampled population. In accord with sound statistical practice, the prior distribution should be examined before sampling to determine the information that might be gained.

Example: To see how Theorem 3 can be applied, suppose we sample servers from a service system at random without replacement and see how many are busy. For an $M/G/N$ loss service system (Riordan 1962, pp. 81–91) the steady-state probability $p_N(k)$ of having k busy servers is characterized by the truncated Poisson distribution, that is, $p_N(k) = x(k)/\sum_{j=0}^{N} x(j)$, $0 \leq k \leq N$, where $x(j) = a_j/j!$. Consequently, $\alpha_N(k) = a/(N - k)$ and $\beta_N(k) = a(N - k - j)/(N - k)$, so that the conditions in (i) and (ii) of Theorem 3 are satisfied. On the other hand, one may easily see that the monotonicity in (i) (or a weaker form) does not extend to all $GI/G/N$ loss systems. For example, in a $GI/G/2$ system (loss or delay) with $P(u > v) = 1$, where u is an interarrival-time random variable and v is a service-time random variable, no more than one server is ever busy.

One may easily see that the strong ordering in Theorem 2 holds in other situations but not necessarily with equal generality. Suppose we have a prior probability mass function $\pi(\theta)$ on a discrete real parameter θ (the discreteness is not necessary); suppose the sample observation or sufficient statistic is real valued and is governed by a conditional probability mass function $f(x|\theta)$. Let $h(\theta|x)$ be the posterior distribution, defined by $h(\theta|x) = f(x|\theta)\pi(\theta)/\sum_\theta f(x|\theta)\pi(\theta)$.

Theorem 4: The posterior mass function $h(\theta|x)$ is MLR nondecreasing (MLR nonincreasing) in x if and only if $f(x|\theta)$ is MLR nondecreasing (MLR nonincreasing) in θ.

Proof: Note that $h(\theta_1|x_2)/h(\theta_1|x_1) \leq h(\theta_2|x_2)/h(\theta_2|x_1)$ if and only if $h(\theta_2|x_1)/h(\theta_1|x_1) \leq h(\theta_2|x_2)/h(\theta_1|x_2)$; however, $h(\theta_2|x)/h(\theta_1|x) = [f(x|\theta_2)/f(x|\theta_1)](\pi(\theta_2)/\pi(\theta_1))$.

Remarks: Theorem 4 covers the example of the beta prior distribution and the binomial sample discussed in the introduction. Theorem 4 is closely related to the theory of uniformly most powerful tests as discussed in Section 5.3 of Ferguson (1967) and references there. Corresponding to Theorem 1 in Ferguson (1967, p. 210), Theorem 4 implies that one-sided tests always minimize Bayes's risk for one-sided hypotheses such as $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$.

[*Received April 1978. Revised October 1978.*]

REFERENCES

DeGroot, Morris H. (1970), *Optimal Statistical Decisions*, New York: McGraw-Hill Book Co.

Ferguson, Thomas S. (1976), *Mathematical Statistics: A Decision Theoretic Approach*, New York: Academic Press.

Kamae, T., Krengel, Ulrich, and O'Brien, George L. (1977), "Stochastic Inequalities on Partially Ordered Spaces," *Annals of Probability*, 5, 899–912.

Riordan, John (1962), *Stochastic Service Systems*, New York: John Wiley & Sons.

Veinott, Arthur, F., Jr. (1965), "Optimal Policy in a Dynamic, Single Product, Nonstationary Inventory Model With Several Demand Classes," *Operations Research*, 13, 761–778.

Whitt, Ward (1980), "Uniform Conditional Stochastic Order," *Journal of Applied Probability*, to appear.