The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1746–1752

THE MAXIMUM VARIANCE OF RESTRICTED UNIMODAL DISTRIBUTIONS

BY HAROLD I. JACOBSON

General Research Corporation

1. Introduction. The least upper bound is derived for the variance of unimodal probability distributions, which are restricted to a finite interval of the real line and possess probability densities with respect to Lebesgue measure. In applications the probability of some rare event is often desired, where the exact form of a distribution and/or its variance are not readily derivable, although the distribution is intuitively known to be unimodal. In such cases an upper bound for the desired probability may be available, e.g., via Chebycheff's inequality, as a function of an upper bound upon the unknown variance. Variance bounds also find use in applications of the Central Limit Theorem. Outside of such applications, it is of separate academic interest to observe the extent to which the condition of unimodality limits the attainable variance.

Johnson and Rogers [2] have shown that for unimodal distributions, the variance is bounded below by $(\text{mean-mode})^2/3$. More recently, Gray and Odell [1] have shown that the variance of certain piecewise continuous functions, restricted to a finite interval, is maximized if taken with respect to the uniform density compared with any other density on the interval that is unimodal, piecewise continuous, and symmetric about the interval midpoint. Their result indicates that the uniform density has the maximum variance within the cited class of symmetric densities.

This paper extends the results of Gray and Odell by dropping the requirement of symmetry–however, at the expense of restricting attention to the distribution variance. Obviously, the amount of available description of the distribution determines the exactitude of variance bounds. Merely knowing that a distribution is restricted to $[a, b]$ serves to bound its variance by $(b - a)^2/4$, which derives from the non–unimodal Bernoulli distribution with atoms of equal probability measure at $x = a$ and $x = b$. It will be shown here that the requirement of unimodality restricts the variance to $(b - a)^2/9$, and that this is a least upper bound. Note that this bound exceeds the variance $(b - a)^2/12$ of a uniform density on $[a, b]$. Some sufficient conditions are also given for the distribution variance not to exceed $(b - a)^2/12$. The moments of the distribution, as with the distribution itself, are presumed unknown.

2. Preliminaries. Let C^* denote the class of probability densities with respect to Lebesgue measure on the real line, that are restricted to a finite interval $[a, b]$, and that are unimodal. Let $C = \{f \varepsilon C^*; \text{some modes of } f \text{ are in the interior } (a, b)\}$. If $x = m$ is a mode of $f(x)$, then $f(x)$ is monotone on $[a, m)$ and on $(m, b]$ and can

Received 25 October 1968.

have only countable discontinuities. Thus we may ignore the indeterminancy of $f(x) \varepsilon C^*$ on its points of discontinuity. In particular, densities are hereby redefined on their discontinuities to ensure the existence of modes, to close the set of modes, and to provide continuity from the left at $x = b$, and from the right at $x = a$. Densities are permitted a value of $+\infty$ at their (single) mode.

We write $E[g(x); f]$ for $\int_a^b g(x)f(x)\,dx$ where it is understood that $f(x)$ is a probability density. $E(f)$ denotes the expectation $E[x; f]$, and variance (f) denotes $E[(x - E(f))^2; f]$. Finally write $M(f; u, v)$ for the mean value of $f(x)$ on the interval (u, v), i.e.,

$$M(f; u, v) = (v - u)^{-1} \int_u^v f(x)\,dx.$$

For a given $f(x) \varepsilon C$ define:

$$t'(f) = \inf \{t_0 \varepsilon [a, b]; \text{ for all } t \geqq t_0, \ f(t) \geqq M(f; t, b)\}$$

$$s'(f) = \sup \{s_0 \varepsilon [a, b]; \text{ for all } s \leqq s_0, \ f(s) \geqq M(f; a, s)\}.$$

(Where the identity of the argument of t' and s' is clear, we write simply t' and s'.) Let $m_1 = \inf\{\text{modes of } f(x)\}$ and $m_2 = \sup\{\text{modes of } f(x)\}$. Then by the unimodality of $f(x)$, $t' \leqq m_1$ and $s' \geqq m_2$. If $m_1 = m_2 = b$, then set $s' = t' = b$. If $m_1 = m_2 = a$, set $s' = t' = a$.

For $f(x) \varepsilon C$ and $t \varepsilon [a, b]$ define:

$$\begin{aligned} f_t(x) &= f(x) & x &< t, \\ &= M(f; t, b) & x &\geqq t. \\ f_s(x) &= M(f; a, s) & x &< s, \\ &= f(x) & x &\geqq s. \end{aligned}$$

Lemma 2.1. *For all $t \varepsilon [t', b]$, $E(f_t)$ is non-increasing in t, and $E(f_t) \rightarrow E(f)$ as $t \rightarrow b$. Also $E(f_s)$ is $\downarrow$ in s for $s \varepsilon [a, s']$, and $E(f_s) \rightarrow E(f)$ as $s \rightarrow a$.*

Proof. For $t' \leqq u < v < b$, the density f_u may be viewed as a modification of f_v obtained by moving probability mass only to the right. Thus $E(f_u) \geqq E(f_v)$, and $E(f_t)$ is $\downarrow$ in t. That $E(f_t) \rightarrow E(f)$ as $t \rightarrow b$, follows because $f(x)$ has no atoms. The statements for $E(f_s)$ follow analogously. □

The point t' (and the analogous s') has the property that for all $t \geqq t'$, there exists in the interval $[t, b]$ only two intervals A and B such that $f(x) > M[f; t, b]$ for $x \varepsilon A$, $f(x) < M[f; t, b]$ for $x \varepsilon B$, and B is to the right of A. Clearly, A is non-empty if and only if B is non-empty. This property of t' and s' is at the heart of Lemma 2.1. Lemma 2.2 may be stated in terms of this property, but we require this lemma in a more general format independent of s' and t'.

Lemma 2.2. *Suppose $f(x) \varepsilon C$ and $g(x) \varepsilon C$ are such that for some $e_j \varepsilon [a, b]$, where $j = 1, \cdots, 8$ and $e_j > e_{j-1}$, one has $f > g$ on (e_1, e_2) and on (e_7, e_8), $f < g$ on (e_3, e_4) and on (e_5, e_6) and $f(x) = g(x)$ otherwise. Then for all $x_0 \varepsilon [e_4, e_5]$, $E[(x - x_0)^2; f] > E[(x - x_0)^2; g]$.*

Lemma 2.2 becomes transparent if $f(x)$ is viewed as a modification of $g(x)$

obtained by moving probability mass strictly away from the point x_0. Its proof is omitted. Note that Lemma 2.2 also holds if $e_j = a$ for $j = 1, \cdots, 4$, or if $e_j = b$ for $j = 5, \cdots, 8$. In these cases probability mass is moved on only one side of the point x_0. Finally, the lemma holds if $e_4 = e_5 = x_0$.

3. Major results. A point $x \varepsilon (a, b)$ will be said to be a point of increase of $f(x)$ if there exists a $\delta > 0$, such that for all u and v in $[a, b]$, $x - \delta < u < x < v < x + \delta$ implies $f(u) < f(v)$. Points of decrease are defined analogously. Consequently, m_1 and m_2 are points of increase and decrease of $f(x)$, respectively, whenever they are in the interior (a, b).

THEOREM 1. *For $f(x) \varepsilon C$ there exists a $g(x) \varepsilon C$ such that*

(i) *for some s^* and t^* with $a \leqq s^* \leqq t^* \leqq b$, and constants k_1, k_2*

$$\begin{aligned} g(x) &= k_1 & x &< s^*, \\ &= k_2 & x &> t^*, \\ &= f(x) & s^* &\leqq x \leqq t^*. \end{aligned}$$

(ii) *If $s^* = t^* = u$, then $E(g) = u$, and $t'(f) \leqq u \leqq s'(f)$.*

(iii) *If $s^* < t^*$ then either:*

(a) *$t^* = E(g)$, $s^* = s'(f)$, and g is $\downarrow$, or*

(b) *$s^* = E(g)$, $t^* = t'(f)$, and g is $\uparrow$.*

(iv) *Variance $(g) \geqq$ variance (f).*

PROOF. Define the function $f_{s,t}(x)$ on [a, b] and for $a \leqq s \leqq t \leqq b$ by:

$$\begin{aligned} f_{s,t}(x) &= M(f; a, s) & x &< s, \\ &= M(f; t, b) & x &> t, \\ &= f(x) & s &\leqq x \leqq t. \end{aligned}$$

Clearly, $f_{s,t} \varepsilon C$ for all s, t, and $E(f_{s,t})$ is a continuous function of s and t on the compact subset $a \leqq s \leqq t \leqq b$ of $[a, b] \times [a, b]$. The intervals in the collection $I = \{[s, t]; s \leqq E(f_{s,t}) \leqq t, s \leqq s'(f), t \geqq t'(f)\}$ are closed subsets of the compact space $[a, b]$. The collection is non-empty because it contains $[a, b]$ In fact I has the finite intersection property, for if not there would be two disjoint intervals $[k, l]$ and $[m, n]$ in I. (This consequence arises from intervals being connected sets of linearly ordered points.) Suppose $l < m$ so that $E(f_{k,l}) < E(f_{m,n})$, $m \leqq s'$, and $l \geqq t'$. But we may apply Lemma 2.1 repeatedly to yield $E(f_{k,l}) \geqq E(f_{k,n}) \geqq E(f_{m,n})$. This contradiction forces the conclusion that $[k, l] \cap [m, n] \neq \emptyset$, and that the collection I has the finite intersection property. Thus for $E_\alpha \varepsilon I$, $\bigcap_\alpha E_\alpha$ is non-empty, and we hereafter denote

$$[s^*, t^*] = \bigcap_\alpha \{E_\alpha ; E_\alpha \varepsilon I\}.$$

If $s^* = t^* = u$, then $E(f_{u,u}) = u$, and $f_{u,u}$ is a step function with a single step. We propose to identify $f_{u,u}(x)$ as the $g(x)$ of the theorem, for $f_{u,u}$ clearly satisfies the first two conditions for $g(x)$. Furthermore, the properties of s' and t' ensure that the intervals $[a, u]$ and $[u, b]$ each contain only two disjoint intervals of the type needed for the application of Lemma 2.2. Lemma 2.2 then yields $E[(x - u)^2; f_{u,u}] > E[(x - u)^2; f]$. This fact and the basic inequality $E[(x - u)^2; f] \geqq E[(x - E(f))^2; f]$ then yield variance $(f_{u,u}) \geqq$ variance (f). Thus, identify $f_{u,u}(x)$ as the desired $g(x)$ of the theorem.

It remains only to consider the case of $s^* < t^*$. In this situation one must have $t^* = t'(f)$ or $s^* = s'(f)$, for if $s^* < s'$ and $t^* > t'$, then it is always possible to find a smaller interval in the collection I by the continuity of $E(f_{s,t})$ in s and t, and by Lemma 2.1. Suppose that $s^* = s'$. Then $m_1 \leqq s' < t^*$ and $f_{s',t^*}(x) = f(x)$ for $x \varepsilon [s', t^*]$, imply that f_{s',t^*} is non-increasing on (s', t^*), and hence on (s', b). Furthermore, for each $\delta > 0$ there exists $x_0 \varepsilon (s', s' + \delta)$ such that $f(x_0) < M(f; a, x_0)$. But $f \downarrow$ on $[s', b]$ implies $M(f; a, x_0) < M(f; a, s')$. Thus the function f_{s',t^*} has a point of decrease at $x = s'$; hence is $\downarrow$ on $[a, b]$. Identify $k_1 = s'$ and $k_2 = t^*$.

It is next claimed that $t^* = E(f_{s',t^*})$, because if $E(f_{s',t}) < t$ one could yet reduce t, hence increase $E(f_{s',t})$ by Lemma 2.1, and by the continuity of $E(f_{s',t})$ in t achieve a state whereby $t = E(f_{s',t})$. Thus, through the intersection over the elements of I we must have $t^* = E(f_{s',t^*})$, and f_{s',t^*} satisfies the third requirement for $g(x)$. Finally, Lemma 2.2 may be applied as before to eventually yield variance $(f) \leqq$ variance (f_{s',t^*}). Thus, set $g(x) = f_{s',t^*}(x)$.

If $t^* = t'$, the function $f_{s^*,t'}(x)$ is everywhere non-decreasing, and may be identified with $g(x)$ through arguments similar to the above. This proves Theorem 1.

THEOREM 2. *For $f(x) \varepsilon C$ there exists a function $h(x) \varepsilon C$ such that:*

(i) *for some constants $d \varepsilon (a, b)$, L_1, and L_2*

$$h(x) = L_1 \qquad x < d,$$
$$= L_2 \qquad x \geqq d;$$

(ii) *if $h(x) \uparrow$ on $[a, b]$, then $(b + a)/2 \leqq E(h) \leqq d$;*

(iii) *if $h(x) \downarrow$ on $[a, b]$, then $d \leqq E(h) \leqq (b + a)/2$;*

(iv) *variance $(h) \geqq$ variance (f).*

PROOF. If in Theorem 1 one has $s^* = t^*$, then the function $g(x)$ of Theorem 1 is identified as the $h(x)$ of Theorem 2. It thus suffices to consider the $g(x)$ of Theorem 1 where $s^* < t^*$, $g(x) \uparrow$ on $[a, b]$, and where $t^* = t'(f)$. For $f(x) \varepsilon$, C, $t'(f) \leqq m_1 < b$ so that $a \leqq s^* < t^* < b$, and $t^* \varepsilon (a, b)$. Hence for each $u \varepsilon [s^*, t^*]$ we may define $h_u(x)$ on $[a, b]$ by

$$h_u(x) = (u - a)^{-1} \int_a^u g(x)\, dx \qquad x < t^*,$$
$$= k_2 + (b - t^*)^{-1} \int_u^{t^*} [g(x) - h_u(t^* - 0)]\, dx \qquad x \geqq t^*.$$

Clearly, $h_u(x) \varepsilon C$ and $E(h_u)$ is a continuous function of u. Lemma 2.1 yields $E(h_{s^*}) \geqq E(g) = s^*$, and $E(h_{t^*}) < E(g) = s^* < t^*$. Thus there exists a $u^* \varepsilon [s^*, t^*)$ such that $u^* = E(h_{u^*})$. We propose to identify $h_{u^*}(x)$ as the $h(x)$ of this theorem.

Clearly, h_{u^*} satisfies the first two conditions for $h(x)$. For condition iv note that h_{u^*} may be characterized as a modification of $g(x)$, obtained by a movement of probability mass strictly away from the point $u^* \varepsilon (s^*, t^*)$. Thus Lemma 2.2 applies to yield $E[(x - u^*)^2; h_{u^*}] \geqq E[(x - u^*)^2; g]$. This result, combined with the basic inequality $E[(x - u^*)^2; g] > E[(x - E(g))^2; g]$ and Theorem 1, yields variance $(h_{u^*}) >$ variance (f). This proves Theorem 2.

Theorem 2 indicates, that in the search for an upper bound on the variance of probability densities in C, it suffices to restrict attention to the functions in C of the form of $h(x)$ in Theorem 2. Such functions will be termed step functions, where it is understood that they possess only a single step. Lemmas 3.1 and 3.2 which follow provide needed characterizations of such step functions.

Lemma 3.1. *Suppose $h(x)$ is a step function in C with step at d, and such that $E(h) = d$. Then, variance $(h) \leqq (b - a)^2/12$, and variance $(h) = (b - a)^2/12$ iff $d = (b + a)/2$.*

Proof. By straightforward calculations one obtains

$$\begin{aligned} h(x) &= (b - d)[(d - a)(b - a)]^{-1} \qquad x < d, \\ &= (d - a)[(b - a)(b - d)]^{-1} \qquad x > d; \end{aligned}$$

and variance $(h) = (b - d)(d - a)/3$. Thus the variance is a quadratic in d, and has a single maximum at $d = (a + b)/2$ where it equals $(b - a)^2/12$. □

Lemma 3.2. *Let $C(d)$ denote the class of step functions in C with step at $d \varepsilon (a, b)$. There exists a member of $C(d)$ of maximum variance. This maximum variance assumes values in $[(b - a)^2/12, (b - a)^2/9]$, and equals $(b - a)^2/12$ iff $d = (b + a)/2$.*

Proof. Let $h(x) \varepsilon C(d)$ equal l for $x > d$. Let $w = l(b - d)(b - a)$. Then variance $(h) = -(w/2)^2 + w[(b - a) + (b - d)]/6 + (d - a)^2/12$. This is a quadratic in w with a single maximum at $w = [(b - a) + (b - d)]/3$, where the variance assumes the value $(b - a)^2(z^2 - z + 1)/9$, with $z = (b - d)/(b - a)$. This last expression is in turn a quadratic in z with a single minimum at $z = \frac{1}{2}$ where the variance equals $(b - a)^2/12$. This maximum variance is a continuous function of z, and as $z \to 1$ or $z \to 0$, the variance approaches and attains the limit $(b - a)^2/9$. □

The following theorem constitutes the major result of this discussion. Recall the definition of C^* given in Section 2.

Theorem 3. *Supremum $\{$variance $(f); f \varepsilon C^*\} = (b - a)^2/9$.*

Proof. First suppose $f(x) \varepsilon C$. Then by Theorem 2, there exists a step function density in C whose variance exceeds variance (f), and whose single step is in the interior of (a, b). Thus by Lemma 3.2, variance $(f) \leqq (b - a)^2/9$.

Next, suppose that $f(x) \varepsilon C^* - C$. It suffices to let the single mode of f be at b. Define

$$\begin{aligned} f_n(x) &= f(x) \qquad\qquad x < b - 1/n, \\ &= f(b - 1/n) \qquad x \geqq b - 1/n; \end{aligned}$$

for integers n exceeding $1/(b - a)$. Then $f_n(x) \uparrow f(x)$ at each $x < b$, and monotone convergence yields $E(x; f_n) \to E(f)$, $E(x^2; f_n) \to E(x^2; f)$, and $k_n \to 1$ where $k_n = 1/E(1; f_n)$ normalizes f_n. Thus on the one hand $k_n f_n \varepsilon C$ and variance $(k_n f_n) \leqq (b - a)^2/9$. On the other hand variance $(k_n f_n) \to$ variance (f). Letting $n \to \infty$ yields variance $(f) \leqq (b - a)^2/9$.

That $(b - a)^2/9$ is a least upper bound for densities in C^* derives from the fact that if $h_d(x) \varepsilon C(d)$ is the density in $C(d)$ of maximum variance (defined by Lemma 3.2), then $h_d \varepsilon C^*$ for all $d \varepsilon (a, b)$, variance (h_d) is a continuous function of d, and variance $(h_d) \to (b - a)^2/9$ as $d \to b$ or as $d \to a$. □

In the proofs of Theorems 1 and 2 cases arose where variance (f) was bounded by the variance of a step function whose step and expectation locations coincided. In such cases Lemma 3.1 yields variance $(f) \leqq (b - a)^2/12$. The following discussion provides some sufficient conditions for this latter variance bound. Let x_0 denote the point $(a + b)/2$.

THEOREM 4. *If* $f(x) \varepsilon C, f(x_0 - 0) \geqq M(f; a, x_0)$, *and* $f(x_0 + 0) \geqq M(f; x_0, b)$, *then variance* $(f) \leqq (b - a)^2/12$.

PROOF. Define $h(x) \varepsilon C$ by

$$h(x) = M(f; a, x_0) \qquad x < x_0,$$
$$= M(f; x_0, b) \qquad x \geqq x_0.$$

The comparison of the densities $h(x)$ and $f(x)$ yields through the agency of Lemma 2.2 that $E[(x - x_0)^2; f] \leqq E[(x - x_0)^2; h]$. This result, taken with $E[(x - E(f))^2; f] \leqq E[(x - x_0)^2; f]$, and the readily verifiable result

$$E[(x - x_0)^2; h] = (b - a)^2/12,$$

together imply variance $(f) \leqq (b - a)^2/12$. □

COROLLARY 4.1. *If* $f(x) \varepsilon C$, *and* f *has a mode at* $(b + a)/2$, *then variance* $(f) \leqq (b - a)^2/12$.

PROOF. $f(x_0)$ exceeds the mean value of $f(x)$ whether taken over $[a, x_0]$ or $[x_0, b]$. □

It follows from Corollary 4.1 that the variance of symmetric densities on $[a, b]$ cannot exceed $(b - a)^2/12$.

COROLLARY 4.2. *If the median of* $f(x)$ *is at* $(a + b)/2$, *then variance* $(f) \leqq (b - a)^2/12$.

PROOF. If $x_0 = (a + b)/2$ is the median of $f(x)$, then the $h(x)$ of Theorem 4 is the rectangular density $h(x) = 1/(b - a)$ with variance $(b - a)^2/12$. Now $f(x)$ must be monotone on either $[a, x_0]$ or $[x_0, b]$. Suppose $f(x)$ is $\uparrow$ on $[a, x_0]$. Then at $x_0 - 0, f(x)$ must exceed or equal its mean value of $1(b - a)$ on $[a, x_0]$. If also at $x_0 + 0, f(x)$ exceeds or equals its mean value of $1/(b - a)$ on (x_0, b), then by Theorem 4 variance $(f) \leqq (b - a)^2/12$. If $f(x_0 + 0) < 1/(b - a)$, then for some $x_1 \varepsilon (x_0, b), f(x_1) > 1/(b - a)$ because $1/(b - a)$ is the mean value of $f(x)$ on (x_0, b). But then $f(x)$ has at least two modes, at $(x_0 - 0)$ and at x_1, so that one must conclude $f(x_0 + 0) \geqq 1(b - a)$. The corollary thus follows by Theorem 4.

REFERENCES

[1] GRAY, H. L. and ODELL, P. L. (1967). On least favorable density functions. *SIAM Rev.* **9** 715–720.

[2] JOHNSON, N. L. and ROGERS, C. A. (1951). The moment problem for unimodal distributions. *Ann. Math. Statist.* **22** 433–439.