Uncertainty, Entropy, Variance and the Effect of Partial Information [1]

by

James V. Zidek [2] and Constance van Eeden

University of British Columbia

Abstract

Uncertainty about the value of an unmeasured real random variable Y is commonly represented by either the entropy or variance of its distribution. If it becomes known that Y lies in a subset A of the support of Y's distribution, one might expect uncertainty about Y to decrease. In other words, one might expect the entropy and variance of Y's conditional distribution given $Y \in A$ to be less than their counterparts for the unconditional distribution. Going further it might be conjectured that the uncertainty about Y would be greater given the knowledge that $Y \in B$ as compared with $Y \in A \subset B$.

We do not know whether these conjectures are correct. However, we give sufficient conditions in certain cases where they are true. In particular, when Y is normally distributed we can make considerable progress. For example, we show in the case that $A = [a, b]$ and Y normally distributed with mean η and variance 1, that the variance of the conditional distribution of Y given that $a \leq Y \leq b$ is less than that of the unconditional distribution, thereby confirming our intuitive reasoning in this case. This last example also shows that for this exponential family the variance is less than 1 for all $a < b$ and all η – a result that is not known among the experts on exponential families we consulted.

[1]Supported by a grant from the Natural Sciences and Engineering Research Council of Canada

[2]corresponding author

The only relevant thing is uncertainty- the extent of our own knowledge and ignorance. (de Finetti, 1970/1974, preface, xi-xii)

1 Introduction.

A search of the Science Citation Index for the years since 1989 lends support to de Finetti's contention. The keyword "uncertainty" yielded 29,386 documents. However, a quick scan of the abstracts for just the most cited documents, as well as standard statistical references reveal a variety of interpretations of this concept, leading us to paraphrase Basu (1975) on "information":

But, what is uncertainty? No other concept in statistics is more elusive in its meaning and less amenable to a generally agreed definition.

As with "information", an agreed on, operationally useful meaning of the term would be desirable. However, such a meaning seems elusive.

Bernardo and Smith (1994) describe uncertainty as "incomplete knowledge in relation to a specified objective". Frey and Rhodes (1996) echo that definition, saying that "uncertainty arises due to a lack of knowledge regarding an unknown quantity". In achieving their comprehensiveness, these definitions suffer from a lack of specificity.

Helton (1997) provides a refinement of the above ideas by classifying "uncertainty" as either stochastic (*i.e.* aleatory) or subjective (*i.e.* epistemic). We have used that classification to help formulate our objectives in this paper.

In an attempt to 'quantify' uncertainty, Frey and Rhodes (1996) propose the use of "probability distributions." O'Hagan (1988) gives a similar interpretation, making "probability" the measure of "uncertainty". Both of these interpretations fail in their attempt to operationalize the concept of uncertainty since the result remains ill-defined except perhaps in the simple case of a single uncertain event, C. [Here, P(C) might be thought of as the "measure" or "quantity" in question. In that case, we might think of $P(C) = 1/2$ as representing the state of maximal uncertainty.] Nevertheless, that simple case and results flowing from it helps us to a more incisive question about the nature of uncertainty and the role of information in connection with it.

Harris (1982) gives a measure of uncertainty that seems to be quite generally accepted, namely the entropy of Y's distribution: $H(Y) = E(-\log f(Y)/m(Y))$ where f denotes the probability density of Y (with respect to counting measure in the discrete case). Here m plays the role of a reference measure against which uncertainty about Y is to be measured. For simplicity, we take $m = 1$ as is commonly done (Singh, 1998, p. 3). Similarly, we define $H(Y|A)$ to be the entropy for the conditional density $f(y|A) = f(y)/P(A), y \in A$ where $P(A) = P(Y \in A)$.

Harris quoting from the celebrated paper of Shannon (1948) states that entropy is "...a measure of choice, uncertainty and information". Renyi (1961) also views the entropy as a measure of both uncertainty and information, the amount of the latter being precisely equal to the amount by which uncertainty will be reduced by the performance of the experiment. The Renyi-Shannon interpretation of entropy as information emphasizes their complementarity. Plausibly, uncertainty should decrease as the amount of information increases. In Section 3, we see that entropy may behave this way in certain cases. However, we are not able to give general answers to even some very simple questions. [Incidentally, the entropy makes $p = P(C) = 1/2$ the maximal state of uncertainty in that it maximizes $-p\log(p) - (1-p)\log(1-p)$, the entropy of the distribution of the indicator random variable, $Z = I_C(Y)$, all in accord with our heuristics, above.]

So in the next section we first study the variance, an even simpler measure of uncertainty. The frequency with which it plays the role of uncertainty, makes it the leading player. In particular, providing standard errors for estimators is often touted as one of the hallmarks of good statistical practice as it indexes the uncertainty attached to the estimate.

The National Institute of Standards and Technology uses the variance or rather, the standard deviation, to express measurement uncertainty. To quote from the Institute's web page (http://physics.nist.gov/cuu/Uncertainty/basic.html) "Each component of uncertainty, however evaluated, is represented by an estimated standard deviation, termed **standard uncertainty** ...".

In one case at least, the variance and entropy are equivalent measures of uncertainty, that case being the normal distribution. Moreover, $p = P(C) = 1/2$ maximizes the variance of Z for the indicator variable, Z defined above, just as it maximized the entropy for the distribution of Z. Thus, these two measures of uncertainty do seem to have some things in common. Since the variance seems simpler than the entropy, we consider it first.

However, even here as we will see in Section 2, answering the questions we ask about the relationship of uncertainty and information proves difficult, even in the normal case. We are able to provide a positive answer to these questions in some cases.

In fact, the genesis of this paper that leads to the issues we address about probability, variance and entropy as measures of aleatory uncertainty is an exceedingly simple question. Would uncertainty about a random variable Y be reduced by the knowledge that $Y \in A$, if $P(A) < 1$? Following the Shannon-Renyi concept of uncertainty- information, we might expect a satisfactory definition of uncertainty would give an affirmative answer to this question. More generally, given two events $A \subset B$, with $P(B - A) > 0$, knowledge that $Y \in A$ should seemingly reduce uncertainty about Y more than $Y \in B$.

The definition proposed by O'Hagan, as implemented here, does not agree with them and shows our heuristics to be too simplistic. More precisely, let $P(Y \in C)$ represent our uncertainty about the event, 'Y falls into C'. Now suppose we learn that Y is in A, for some A for which $A \cap C$ is not empty. Will our uncertainty about the event decrease as our heuristics suggest?

The answer to that question can be 'no' when the new knowledge conflicts with our prior beliefs. The new probability of the event, $P(Y \in C|Y \in A)$, may be closer to 1/2 than $P(Y \in C)$. For example, if $Y \sim U[0,1]$, C=[0,c], and A=[a,1] with $0 < a < c < 1$, then $P(Y \in C|Y \in A) = (c-a)/(1-a) = 1/2$ when a= 2c-1. If c=7/8 and a =3/4 we would move from a state of near certainty about the event to one of complete uncertainty (conditional probability 1/2) about whether or not it has occurred. New information need not reduce our uncertainty with this simple measure of uncertainty. Can the same phenomenon occur with other measures of uncertainty? That question is the subject of the next two sections.

2 Variance.

In this section, we study the behaviour of the variance of a real random variable Y as a measure of uncertainty.
In particular, we consider what happens to that measure when either the knowledge that $Y \in B$ or $Y \in A = [a, b] \subset B$ obtains.

We first prove, for $A = [-c, c]$, a rather general result that falls in line with our intuition.

Theorem 2.1 *Let Y be a real random variable symmetrically distributed around zero, let $I = \{\ c \mid P(|Y| \leq c) > 0\}$, and let $A_c = [-c, c]$ for $c \in I$. Then $V(Y|A_c)$ is nondecreasing in c. Further, for $c \in I$ and $c < c_1$, $V(Y|A_c) = V(Y|A_{c_1})$ if an only if $P(c < Y \leq c_1) = 0$.*

Proof. Let $\mathcal{G}$ be the distribution function of Y^2. Then, with $\gamma = c^2$ and $\gamma_1 = c_1^2$, $V(Y|A_{c_1}) - V(Y|A_c)$ has the same sign as

$$\begin{aligned}\int_{[0,\gamma_1]} s d\mathcal{G}(s) \int_{[0,\gamma]} d\mathcal{G}(t) - \int_{[0,\gamma_1]} d\mathcal{G}(s) \int_{[0,\gamma]} t d\mathcal{G}(t) &= \int_{[0,\gamma_1]} \int_{[0,\gamma]} (s-t) d\mathcal{G}(s) d\mathcal{G}(t) \\ &= \int_{(\gamma,\gamma_1]} \int_{[0,\gamma]} (s-t) d\mathcal{G}(s) d\mathcal{G}(t) \\ &\geq 0.\end{aligned}$$

Further, the above inequality is an equality if and only if $\int_{(\gamma,\gamma_1]} \int_{[0,\gamma]} d\mathcal{G}(s) d\mathcal{G}(t) = P(c < Y \leq c_1) P(0 \leq Y \leq c_1) = 0$. Since $c_1 \in I$, that latter is equivalent to $P(c < Y \leq c_1) = 0$.
□

Although this result seems extremely general, we have imposed an important restriction, that Y is symmetrically distributed around zero. That makes the new information, 'Y is in A', confirmatory, the more so when c is small. Therefore, it does not seem surprising that the new information has made the level of uncertainty grow smaller. To probe further entails looking for information that is inconsistent with our prior model. Thus, we keep A symmetric as above, but make the distribution asymmetric about 0 so that, in particular, its mean is no longer 0. We have chosen $Y \sim N(\eta, 1)$, because of the central role of the normal distribution in statistical theory. What we can then show as the next result is that the result in the previous theorem is true when the event $B = (-\infty, \infty)$ and η is arbitrary.

Theorem 2.2 $V(Y|A) < V(Y) = 1$ *for all* η *when* $A = [-c, c]$.

Proof. Using the fact that the conditional distribution of Y given that $Y \in A$ is an exponential family distribution, we have

$$V(Y|A) = \frac{d}{d\eta}\mathcal{E}(Y|A),$$

where

$$\mathcal{E}(Y|A) = \frac{\int_{-c}^{c} y e^{-(y-\eta)^2/2} dy}{\int_{-c}^{c} e^{-(y-\eta)^2/2} dy} = \eta - \frac{\phi(\eta - c) - \phi(\eta + c)}{\Phi(\eta + c) - \Phi(\eta - c)},$$

where ϕ and Φ are, respectively, the standard normal density and distribution function. So, $V(Y|A)$ can be written as

$$V(Y|A) = 1 - \frac{d}{d\eta}\frac{\phi(\eta - c) - \phi(\eta + c)}{\Phi(\eta + c) - \Phi(\eta - c)}$$

and it is now sufficient to show that

$$h(\eta) = \frac{\phi(\eta - c) - \phi(\eta + c)}{\Phi(\eta + c) - \Phi(\eta - c)} \tag{2.1}$$

is increasing in η. Given that $h(\eta) = -h(-\eta)$ for all η, it is sufficient to prove this for $\eta \geq 0$. A proof of this result can be found in van Eeden and Zidek (2001) but, given that it is very short, we repeat it here for completeness. Note that

$$h'(\eta) = K(\eta)[(\eta + c)\phi(\eta + c) - (\eta - c)\phi(\eta - c) + h(\eta)(\phi(\eta - c) - \phi(\eta + c))]$$

where $K(\eta)^{-1} = \Phi(\eta + c) - \Phi(\eta - c)$. Clearly $h'(\eta) > 0$ when $0 \leq \eta \leq c$, because $\phi(\eta - c) > \phi(\eta + c)$ when $\eta \geq 0$. For $\eta > c$

$$K^{-1}(\eta) = \int_{\eta - c}^{\eta + c} \phi(t)dt < \int_{\eta - c}^{\eta + c} \frac{t}{\eta - c}\phi(t)dt = \frac{\phi(\eta - c) - \phi(\eta + c)}{\eta - c},$$

which gives $h'(\eta) > 2cK(\eta)\phi(\eta + c) > 0$. The result then follows from the fact that h is continuous. □

Remark 2.1 As an anonymous referee observed, the previous theorem can easily be extended to the case where $A = [a, b]$. We may simply transform Y into $Y' = Y - (b + a)/2$ so that η gets transformed into $\eta' = \eta - (b + a)/2$. We may now apply the theorem to obtain the conclusion of the theorem.

Remark 2.2 The same referee also obtained the result in the last theorem for an infinite interval, say $[a, \infty)$. The proof for this case differs from that which we give for the finite case, but for brevity is omitted.

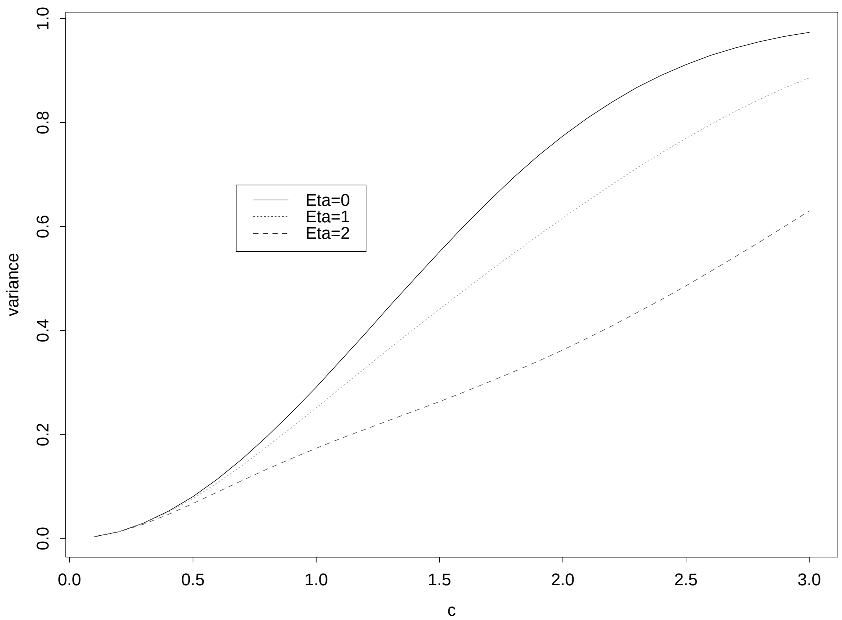

Figure 1: Conditional Variance of Y as a Function of c for Various Values of Eta.

The heuristics given in Section 1 would lead us to conjecture that $V(Y|A)$ is increasing in c when $A = [-c, c]$. That conjecture seems to be supported by the plots of the conditional variance against c for various values of η depicted in Figure 1. However, we have not been able to prove or disprove this conjecture other than in the case considered in Theorem 2.1. But we have necessary and sufficient conditions for the derivative of $V(Y|A)$ with respect to c to be positive under very minimal conditions on the density f. They are given in Theorem 2.3.

Theorem 2.3 *If $A = [-c, c]$ and Y has density $f(y)$ which is positive for $y \in (-\infty, \infty)$, then $(d/dc)V(Y|A) > 0$ if and only if*

$$c^2 - \mathcal{E}(Y^2|A) + 2\mathcal{E}(Y|A)\left(\mathcal{E}(Y|A) - c\,\frac{f(c) - f(-c)}{f(c) + f(-c)}\right) > 0.$$

Proof. The conditional variance of Y given that $Y \in [-c, c]$ is

$$V(Y|A) = \frac{1}{P(A)} \int_{-c}^{c} y^2 f(y)dy - \left(\frac{1}{P(A)} \int_{-c}^{c} y f(y)dy\right)^2.$$

So

$$\frac{d}{dc}V(Y|A) = \frac{c^2(f(c) + f(-c))}{P(A)} - \frac{f(c) + f(-c)}{(P(A))^2} \int_{-c}^{c} y^2 f(y)dy$$

$$-\frac{2c(f(c) - f(-c))}{(P(A))^2} \int_{-c}^{c} y f(y)dy + \frac{2(f(c) + f(-c))}{(P(A))^3} \left(\int_{-c}^{c} y f(y)dy\right)^2 =$$

$$\frac{f(c)+f(-c)}{P(A)}\left(c^2-\mathcal{E}(Y^2|A)+2\left(\mathcal{E}Y|A)\right)^2-2c\mathcal{E}(Y|A)\frac{f(c)-f(-c)}{f(c)+f(-c)}\right).$$

The result then follows from the fact that $f(y)>0$ for $y \in (-\infty, \infty)$. □

Following the reading of this paper, Dr Chris Klaassen, proved a theorem with the same conclusion as Theorem 2.1 but under very different hypotheses. He has graciously allowed us to present his result as Theorem 2.4. As in our theorem, the conditional variance is shown to be monotone while the distribution is symmetric about 0. However, unlike our theorem and more like our Theorem 2.2, the new information in his theorem tends to be inconsistent with the prior model.

Theorem 2.4 *(Klaassen) Let Y be symmetric about zero with distribution function F and let C be the smallest real such that F puts all its mass on $[-C, C]$. Let V_c be the conditional variance of Y given that $|Y| > c$. If $\mathcal{E}Y^2$ is finite, then V_c is a nondecreasing function of $c \in [0, C)$ and strictly increasing at all such c at which F is strictly monotone.*

Proof. First note that

$$\begin{aligned}
V_c &= \frac{\int_{c+}^{C} y^2 dF(y)}{\int_{c+}^{C} dF(y)} \\
&= \frac{[y^2(F(y)-1)]_{c+}^{C} + 2\int_{c+}^{C} y(1-F(y))dy}{1-F(c)} \\
&= c^2 + \frac{2}{1-F(c)}\int_{c+}^{C} y(1-F(y))dy.
\end{aligned}$$

Now let, for c and d in $[0, C)$

$$\chi(c,d) = c^2 + \frac{2}{1-F(d)}\int_{c+}^{C} y(1-F(y))dy,$$

then $\chi(c,c) = V_c$ and $\chi(c,d)$ is nondecreasing in d for all c. Further, for any $0 \leq d \leq c < C$,

$$\frac{d}{dc}\chi(c,d) = 2c\frac{F(c)-F(d)}{1-F(d)} \geq 0.$$

Therefore we have, for $0 < \varepsilon < C - c$,

$$V_{c+\varepsilon} = \chi_{c+\varepsilon,c+\varepsilon} \geq \chi_{c+\varepsilon,c} \geq \chi_{c,c} = V_c. \tag{2.2}$$

Further note that the first inequality in (2.2) is strict when $F(c+\varepsilon) > F(c)$. □

3 The Entropy.

To begin, we state an easily proven identity using $\bar{A}$ denote the complement of A. With this notation, we have

$$H(Y) = P(A)H(Y|A) + P(\bar{A})H(Y|\bar{A}) + H(I_A\{Y\}) \tag{3.1}$$

where I_A is the indicator function for the set A. Therefore, $H(Y) - H(Y|A) = P(\bar{A})[H(Y|\bar{A}) - H(Y|A)] + H(I_A\{Y\})$.

From this identity we can see that knowing $Y \in A$ will reduce our uncertainty about Y (regardless of our choice of the reference measure, m, above) if:

- our uncertainty about whether or not Y will lie in A and hence the last term in identity (3.1) is large;
- knowing that Y is in A makes us appreciably more certain about the value of Y than does knowing that Y is in $\bar{A}$.
- $Y \in \bar{A}$ is very unlikely

Otherwise, our uncertainty about Y might increase as a result of partial knowledge about Y thus flying in the face of our intuition.

To shed some light on this issue, we examine the identity above more closely and see that the unconditional entropy is sensitive to what might be called 'model uncertainty' as well as uncertainty about Y itself. That model is either $f(y|A)$ or $f(y|\bar{A})$ as appropriate. Moreover, if $Y \in A$ is in disagreement with the model $f(y|A)$, that new knowledge can leave us more uncertain about Y than before. However, we have not been able to find an example for the entropy where this phenomenon actually arises.

The next theorem gives conditions that ensure the monotonicity of entropy suggested by the heuristics in Section 1.

Theorem 3.1 *Suppose $A \subset B$, $P(A) > 0$ and $\inf_{y\in A} f(y)/\sup_{y\in B-A} f(y) \geq e^{-1}$ where f denotes the density function of a random variable Y. Then $H(Y|A) \leq H(Y|B)$.*

Proof. The result is obviously true when $P(B-A) = 0$. So now suppose that $P(B-A) > 0$. Then

$$\begin{aligned} H(Y|B) - H(Y|A) &= \frac{1}{P(B)}\int_B(-\log f(y))f(y)dy - \frac{1}{P(A)}\int_A(-\log f(y))f(y)dy \\ &+ \log P(B) - \log P(A), \end{aligned}$$

where

$$\log P(B) - \log P(A) = \int_{P(A)}^{P(B)} \frac{1}{u}\, du \geq \frac{P(B) - P(A)}{P(B)} = \frac{P(B-A)}{P(B)}.$$

Thus

$$\frac{P(B)[H(Y|B) - H(Y|A)]}{P(B-A)} \geq$$

$$\frac{\int_{B-A}(-\log f(y))f(y)dy}{P(B-A)} + \int_A (\log f(y))f(y)dy \left[\frac{P(B)}{P(A)} - 1\right][P(B-A)]^{-1} + 1 =$$

$$\frac{1}{P(B-A)}\int_{B-A}(-\log f(y) + 1)f(y)dy + \frac{1}{P(A)}\int_A (\log f(y)f(y)dy \geq$$

$$\inf_{y\in B-A}(1 - \log f(y)) + \inf_{y\in A}(\log f(y)) = 1 - \log(\sup_{y\in B-A} f(y)) + \log(\inf_{y\in A} f(y)) \geq 0$$

if $inf_{y\in A} f(y)/\sup_{y\in B-A} f(y) \geq e^{-1}$. □

Observe in the previous theorem that we do not require Y to be real valued.

Example. Suppose $Y \sim N_p(0, \Sigma)$, while $A = \{y : ||y|| \leq c\}$ and $B = \{y : ||y|| \leq d\}$ where $c < d$. Then it is easily seen by the previous result that $H(Y|A) \leq H(Y|B)$.

Two corollaries follow from the previous theorem. The first concerns densities like that of the normal that have the shape of an "inverted bowl."

Corollary 3.1 *Suppose that $A = [-c, c]$, $B = [-d, d]$ with $c < d$, that f is symmetric about 0 and that f is non-increasing for $y > 0$. Then $H(Y|A) \leq H(Y|B)$.*

Example. Suppose Y has a Cauchy distribution centered at $y = 0$ and the sets A and B are as in the previous corollary. Then $H(Y|A) \leq H(Y|B)$.

Corollary 3.2 *Suppose $A = [c, d]$ while $B = [b, d]$ where $b < c < d \leq b + \sqrt{2}$ and that $Y \sim N(\eta, 1)$. Then $H(Y|A) \leq H(Y|B)$ provided that $(b + d)/2 - 1/(d - b) \leq \eta$.*

Proof. Four distinct cases need to be considered, namely 1) $\eta < b$; 2) $b \leq \eta < c$; 3) $c \leq \eta < (c + d)/2$; 4) $\eta \geq (c + d)/2$.
In case 4), where $\eta \geq (c + d)/2$, $-\log(\sup_{y\in B-A} f(y)) + \log(\inf_{y\in A} f(y)) = 0$, because the sup and inf are both attained at $y = c$. So, $H(Y|B) - H(Y|A) \geq 0$ in case 4), whatever be $b < c < d$.

In case 3), where $c \le \eta < (c+d)/2$,

$$1 - \log(\sup_{y \in B-A} f(y)) + \log \inf_{y \in A} f(y)) = 1 + \frac{1}{2}(c-\eta)^2 - \frac{1}{2}(d-\eta)^2,$$

which is nonnegative if and only if

$$(d-c)\left(\frac{d+c}{2} - \eta\right) \le 1.$$

Combining this with $c \le \eta < (c+d)/2$ gives the condition

$$c = \max\left(c, \frac{d+c}{2} - \frac{1}{d-c}\right) \le \eta < \frac{c+d}{2}.$$

So, (c,d) has to satisfy $c \ge (d+c)/2 - (1/(d-c))$, i.e. $d \le c + \sqrt{2}$, which is implied by $d \le b + \sqrt{2}$.

Now consider case 2), where $b \le \eta < c$. Here

$$1 - \log_{y \in B-A} \sup(f(y)) + \log_{y \in A} f(y) = 1 - \frac{1}{2}(d-\eta)^2.$$

So, here we need

$$b = \max(b, d - \sqrt{2}) < c,$$

which is satisfied when $d \le b + \sqrt{2}$.

Finally, for case 1) where $\eta < b$,

$$1 - \log_{y \in B-A} \sup(f(y)) + \log_{y \in A} f(y) =$$

$$1 + \frac{1}{2}(b-\eta)^2 - \frac{1}{2}(d-\eta)^2 = 1 + (b-d)\frac{b+d}{2} + (d-b)\eta,$$

which is nonnegative if and only if $\eta \ge (b+d)/2 - 1/(d-b)$, a condition we imposed. Further $(b+d)/2 - 1/(d-b) \le b$ is satisfied when $d \le b + \sqrt{2}$. □

We complete this section by proving a theorem for the normal distribution where η is not restricted. However, to gain that generality, we are limited by our method of proof to the case of $B = (-\infty, \infty)$.

Theorem 3.2 *Suppose $A = [-c, c]$ and $Y \sim N(\eta, 1)$. Then, when $B = (-\infty, \infty)$, $H(Y|A) \le H(Y|B) = H(Y)$ for all η.*

Proof. For a $\mathcal{N}(\eta, 1)$ distribution

$$H(Y) = \mathcal{E}(-\log f(Y)) = \log(\sqrt{2\pi}) + \frac{1}{2}\mathcal{E}(Y-\eta)^2 = \frac{1}{2}(\log(2\pi) + 1).$$

Further, $f(y|A) = f(y)/P(A)$ which gives

$$H(Y|A) = \mathcal{E}(-\log f(Y|A)|A) = \mathcal{E}\left(\log(\sqrt{2\pi}) + \frac{1}{2}(Y-\eta)^2 + \log(P(A) \mid A\right) =$$

$$\log\sqrt{2\pi} + \log P(A) + \frac{1}{2}V(Y|A) + \frac{1}{2}(\mathcal{E}(Y|A)-\eta)^2 =$$

$$H(Y) + \frac{1}{2}\left(2\log P(A) + V(Y|A) + (\mathcal{E}(Y|A)-\eta)^2 - 1\right),$$

where $V(Y|A)$ is the conditional, given $Y \in A$, variance of Y.
So, $H(Y|A) - H(Y) < 0$ if and only if

$$-\log P(A) - \frac{1}{2}(V(Y|A)-1) - \frac{1}{2}(\mathcal{E}(Y|A)-\eta)^2 > 0 \quad \text{for } \eta > 0, \tag{3.2}$$

From the proof of Theorem 2.2 we know that $\mathcal{E}(Y|A) = \eta - h(\eta)$ and $1 - V(Y|A) = h'(\eta)$, where $h(\eta)$ is given by (2.1). This shows that (3.2) is equivalent to

$$-\log P(A) + \frac{1}{2}h'(\eta) - \frac{1}{2}h^2(\eta) > 0 \quad \text{for } \eta > 0.$$

So, now it is sufficient to show that

$$G(\eta) = -\log P(A) - \frac{1}{2}h^2(\eta) > 0 \quad \text{for } \eta > 0,$$

because (see the proof of Theorem 2.2) $h'(\eta) > 0$ for all η. But $G(0) = -\log(\Phi(c) - \Phi(-c)) > 0$ because $h(0) = 0$. Further,

$$\begin{aligned} \frac{d}{d\eta}G(\eta) &= -\frac{\phi(\eta+c)-\phi(\eta-c)}{\Phi(\eta+c)-\Phi(\eta-c)} - h(\eta)h'(\eta) \\ &= h(\eta) - h(\eta)h'(\eta) = h(\eta)V(Y|A) > 0 \quad \text{for } \eta > 0, \end{aligned}$$

which proves the result. □

4 Concluding Remarks.

Statistics is rooted in uncertainty. Yet despite its importance, few measures of that quantity have been proposed in contrast to its cousin, "information". That seems especially surprising in light of the knowledge that at least for aleatory uncertainty, they would seem to be complementary concepts.

In this paper, we have studied two of the most common measures seen in statistical contexts, the variance and the entropy. We investigate how they behave when partial information about a random variable Y becomes available namely, that Y belongs to a subset A.

Although the subject of our investigation seems simple, we have not been able to give any general conclusions. For example, we cannot assert that the residual uncertainty about Y given $Y \in A$ "increases" as A increases, as one might naively expect. Instead, we have obtained results for a number of special cases, leaving many questions for future study.

Acknowledgements. We are indebted to an anonymous referee for the suggestions referred to in Remarks 2.1 and 2.2. We also thank the Editor for additional suggestions and his expeditious handling of this manuscript. Finally, we are extremely grateful to Chris Klaassen for stimulating conversations on the topic of this paper and for allowing us to include his Theorem 2.4.

Bibliography

BASU, D. (1975). Statistical information and likelihood. *Sankhyā, Ser A*, 37, 1-71.

BERNARDO, J.M. AND SMITH, A.F.M. (1994). *Bayesian Theory.* Chichester: Wiley.

DE FINETTI, B. (1970/1974). *Teoria delle Probabilità* 1, Turin: Einaudi. English translation as *Theory of Probability* 1 in 1974, Chichester, Wiley.

FREY, H.C. AND RHODES, D.S. (1996). Characterizing, simulating, and analyzing variability and uncertainty: An illustration of methods using an air toxics emissions example. *Human and Ecological Risk Assessment*, 2, 762-797.

HARRIS, B. (1982). Entropy. In *Encyclopedia of Statistical Science, Vol 2.* Eds S. Kotz and N.L. Johnson, 512-516.

HELTON, J.C. (1997). Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. *J. Statist. Comput. Simulation*, 57, 3-76.

O'HAGAN, A. (1988). *Probability: Methods and Measurement.* London:Chapman and Hall.

RÉNYI, A. (1961). On measures of entropy and information. In *Proceedings Fourth Berkeley Symposium*, Ed. J. Neyman, 547-561.

SHANNON, C.E. (1948). A mathematical theory of communication. *Bell System Tech. J.*, 27, 379-423, 623-656.

SINGH, V.P. (1998). *Entropy-based parameter estimation in hydrology.* Dordrecht: Kluwer.

van Eeden, C. and Zidek, J.V. (2001). Estimating one of two normal means when their difference is bounded. *Statist. Probab. Lett.*, 51, 277-284.