Uncertainty of Random Variables

Fabián, Zdeněk
2010

Institute of Computer Science
Academy of Sciences of the Czech Republic

Uncertainty of random variables

Zdeněk Fabián

Technical report No. 1084

September 2010

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: +420 266 053 210, fax: +420 286 585 789,
e-mail:zdenek@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Uncertainty of random variables[1]

Zdeněk Fabián

Technical report No. 1084

September 2010

Abstract:

New characteristics of continuous random variables introduced in [4]-[6] are generalized for discrete random variables. It makes possible to introduce uncertainty function of random variable and compare its mean value with the Shannon entropy.

Keywords:
scalar score, information, entropy

[1]The research presented in the paper was supported by grant ME 949 "Analysis of negative impacts on driver attention" of The Ministry of Education, Youth and Sports of the Czech Republic.

1 Introduction

Let X be a discrete random variable with probability mass function $f(k), k = 0, 1, ..., n$. From the times of the Shannon's discovery, the uncertainty of X before the experiment or information contained in a realization x of X after the experiment is expressed, for any k, as $U(k) = \log(1/f(k))$. If the result of an experiment is more or less expected, uncertainty is low, whereas an unexpected result with low probability $f(k)$ carries a great amount of uncertainty. The mean value of this "uncertainty function" is the entropy

$$H(X) = \mathrm{E}U(k) = \sum_{k=0}^{n} \log(1/f(k))\, f(k) = \sum_{k=0}^{n} -f(k)\log f(k). \tag{1.1}$$

Let X be a continuous random variable with support set $\mathcal{X} = (a, b) \subseteq \mathbb{R}$, distribution F and density f. The analogy of the Shannon entropy for continuous random variables is the differential entropy

$$h(X) = \mathrm{E}\log(1/f(x)) = \int_{\mathcal{X}} -\log f(x) f(x)\; dx. \tag{1.2}$$

Since $U(x) = \log(1/f(x))$ can be negative in certain range of parameters practically for any parametric distribution, it can be hardly considered to be an "uncertainty function". Even the mean value EU can be negative, too. This is the reason that statisticians prefer the Fisher information. However, Fisher information relates to the parameters of parametric distributions. The generalization presented in [2] is meaningful for distributions with support $\mathbb{R}$ only.

In [3]-[5] we introduced to a given regular continuous distribution F a scalar function $S(x)$, called now the scalar score. It appeared that $S^2(x)$ can be considered as expressing information contained in observation x in the given model. In the present paper we briefly describe the scalar score and introduce an uncertainty function based on $S^2(x)$. In the last section we generalize concept of the scalar score for discrete distributions and show a relation between the mean uncertainty and the Shannon entropy.

2 Scalar score

As an important function of a distribution G with support $\mathbb{R}$ and density g we identified, using lesson drawn from [6], the *score function*

$$T_G(y) = -\frac{1}{g(y)}\frac{d}{dy}g(y). \tag{2.1}$$

Value ET_G^2 is the generalized Fisher information introduced in [2].

Let $\eta : \mathcal{X} \to \mathbb{R}$ be a suitable mapping. As an important function describing the transformed distribution on $\mathcal{X} \neq \mathbb{R}$,

$$F(x) = G(\eta(x)), \quad x \in \mathcal{X}, \tag{2.2}$$

was suggested in [4] the transformed score function of G,

$$T(x) = T_G(\eta(x)). \tag{2.3}$$

From (2.1) and (2.2) we obtain

$$T(x) = -\frac{1}{f(x)}\frac{d}{dx}\left(\frac{1}{\eta'(x)}f(x)\right), \tag{2.4}$$

where $\eta'(x) = d\eta(x)/dx$ is the Jacobian of the transformation.

For a comparison of properties of function (2.4) of different distributions, it turned out to be necessary to use one concrete $\eta : \mathcal{X} \to \mathbb{R}$ for all distributions with a given support. According the principle of parsimony, that one providing the simplest mathematical forms of (2.4) for a large amount of commonly used distributions should be used. According [7] and [5], η was defined as

$$\eta(x) = \begin{cases} x & \text{if} \quad \mathcal{X} = \mathbb{R} \\ \log(x-a) & \text{if} \quad \mathcal{X} = (a,\infty) \\ \log\dfrac{(x-a)}{(b-x)} & \text{if} \quad \mathcal{X} = (a,b). \end{cases} \tag{2.5}$$

Function (2.4) with η given by (2.5) is called the *transformation-based score* or shortly the t-score.

Under mild regularity condition, the transformation-based score is a unique description of distributions, expressing the relative change of a "basic component" of the density of the model (the density divided by Jacobian of the transformation) with respect to the probability density.

T-scores of some distributions are well-known functions. The t-score of the standard normal distribution is $T(x) = x$. The t-score of a location distribution with support $\mathbb{R}$ and location parameter μ (expressing the location of the maximum of the density) is the score function

$$T_G(y-\mu) = \frac{\partial}{\partial\mu}\log g(y-\mu). \tag{2.6}$$

The log-location distributions [8] are distributions transformed from $\mathbb{R}$ into $\mathcal{X} = (0,\infty)$ by $\eta(x) = \log(x)$ with "transformed location" parameter $\tau = \exp(\mu)$. By [4], Theorem 1, it holds for them that

$$S(x;\tau) \equiv \eta'(\tau)T(x;\tau) = \frac{\partial}{\partial\tau}\log f(x;\tau), \tag{2.7}$$

which is the likelihood score for τ.

It is easy to see using (2.6) and (2.3) that $T(\tau;\tau) = 0$. Moreover, the value $ES^2 = \int_{\mathcal{X}} S(x;\tau)^2 f(x)\,dx$ is the Fisher information for τ.

Our basic notions, parameter τ and inference function S of log-location distributions, were generalized for arbitrary distribution as follows:

As the most important point of the distribution, expressing its central tendency, was identified, instead of τ, the zero of the t-score, the solution x^* of equation

$$T(x) = 0,$$

called the *t-mean*. The t-mean is actually the transformed mode (the maximum of the density) of the prototype. It is an easily manipulated number which is not far from the mean of light-tailed distributions, being a reasonable "center" of heavy-tailed and skewed distributions.

Function (2.7) was generalized by using the t-mean instead of τ by

$$S(x) = \eta'(x^*)T(x). \tag{2.8}$$

We call it a *scalar score of distribution F*. Scalar scores of parametric distributions $S(x;\theta) = \eta'(x^*)T(x;\theta)$ were suggested as inference functions for adapting the data to the assumed parametric model. For a given x, $S(x)$ describes the sensitivity of the t-mean to the value x. Function $S(x,\theta)$ as a function of θ is the "likelihood score for x^*" either x^* is a parameter of the distribution or not.

The sample mean and sample variance of distributions with probability densities approaching to zero too slowly (the heavy tailed distributions) are not relevant characteristics of the data since the integrals defining the moments can be infinite. It follows from (2.1) that if $g(y) = O(e^{-y})$ if $y \to \pm\infty$ then $T_G(y) = O(1)$. Since (2.5) retains the properties of t-scores on boundaries of the support, the scalar scores of heavy-tailed distributions are bounded.

Function $S^2(x)$ attains its minimum at x^* (proof: the density of (2.2) is $f(x) = g(\eta(x))\eta'(x)$. The term $\eta'(x)$ is common to all distributions with the given support and does not carry any information about X. The first term is minimal if $\frac{d}{dx}g(\eta(x)) = \frac{d}{dx}(f(x)/\eta'(x)) = 0$, which gives $T(x) = 0$ by (2.4)). Further, $S^2(x)$ increases from x^* quickly/slowly if S is unbounded/bounded. Under the usual regularity conditions $\mathrm{E}S^2$ is finite and means information. We thus insist that function $S^2(x)$ could play the role of the Fisher information function of continuous random variables, giving a relative

Distribution	exponential	gamma	Weibull	lognormal
σ^2	τ^2	α/γ^2	$\frac{\pi^2}{6}\tau^2 s^2$	$\tau^2 e^{s^2}(e^{s^2}-1)$
ω^2	τ^2	α/γ^2	$\tau^2 s^2$	$\tau^2 s^2$

Table 2.1: Ordinary and score variance of light-tailed distributions

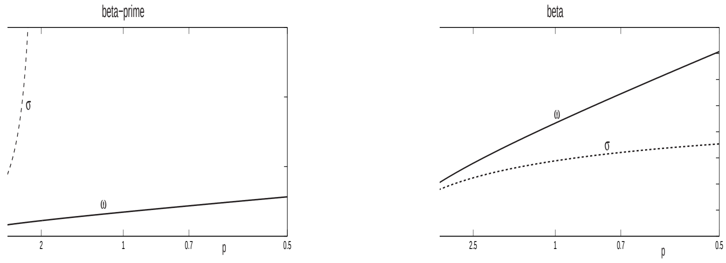

Figure 2.1: σ and ω of the beta-prime and beta distributions as functions of p.

information contained in observation x, small if the distribution is heavy-tailed,being vast if x is an outlier in a model which not "expect" an occurrence of outlying observations.

Since $\eta'(x^*) \neq 0$ and $\mathrm{E}S^2 > 0$, the *score variance*

$$\omega^2 = \frac{1}{\mathrm{E}S^2} = \frac{1}{[\eta'(x^*)]^2 ET^2(\theta)} \tag{2.9}$$

is finite. The score variance of distributions with $\mathcal{X} = (0, \infty)$, $\omega^2 = (x^*)^2/ET^2$, is proportional to the square of the t-mean, which is in agreement with σ^2 of light-tailed distributions (see Table 2, where we denoted $s = 1/c$. The value σ^2 of the Weibull distribution is an approximation for low s).

The score variance of heavy-tailed distributions, however, is a new quantity. The left panel of Fig. 1 compares ω and σ of the beta-prime distribution for $q = p$, where ω^2 is given in Table 2 below and $\sigma^2 = \frac{p(p+1)}{(q-1)(q-2))}$. The ordinary σ blows up at $q = 2$.

The score variance of distributions with support $\mathcal{X} = (-b, b)$ is $\omega^2 = \frac{b^2}{16ET^2}$. For the uniform distribution with $f(x) = \frac{1}{2b}$ thus $\omega^2 = \frac{3b^2}{4}$, whereas the ordinary $\sigma^2 = \frac{b^2}{3}$. The right panel of Fig. 1 shows σ and $\omega = (\frac{2p+1}{p^2})^{1/2}$ of the beta distribution, $q = p$. Measure ω assigns large values to U-shaped distributions with $p < 1$.

3 Uncertainty function

Definition 1. Let X be random variable with distribution F with support set $\mathcal{X}$. Denote by f its density, T the t-score and x^* the t-mean. Let η be given by (2.5) and S be the scalar score given by (2.8). Define the uncertainty function of X by

$$U(x) = \frac{S^2(x)}{(\mathrm{E}S^2)^2}. \tag{3.1}$$

The uncertainty function is defined by such a way that the mean uncertainty equals to the score variance, $EU = \omega^2$. $U(x)$ can be determined also from relation

$$U(x) = \omega^2 \frac{T^2(x)}{ET^2}, \tag{3.2}$$

equivalent to (3.1). Table 2 and Fig. 2 shows uncertainty functions of some currently used distributions.

F	$\mathcal{X}$	$f(x)$	$T(x)$	ω^2	$U(x)$
normal	R	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$\frac{1}{\sigma}\frac{x-\mu}{\sigma}$	σ^2	$(x-\mu)^2$
Cauchy	R	$\frac{1}{\pi\sigma(1+(\frac{x-\mu}{\sigma})^2)}$	$\frac{1}{\sigma}\frac{2\frac{x-\mu}{\sigma}}{1+(\frac{x-\mu}{\sigma})^2}$	$2\sigma^2$	$\frac{16(x-\mu)^2}{\left(1+(\frac{x-\mu}{\sigma})^2\right)^2}$
lognormal	$(0,\infty)$	$\frac{c}{\sqrt{2\pi}x}e^{-\frac{1}{2}\log^2(\frac{x}{\tau})^c}$	$c\log(\frac{x}{\tau})^c$	$\frac{\tau^2}{c^2}$	$\frac{\tau^2}{c^2}\log^2(\frac{x}{\tau})^c$
Weibull	$(0,\infty)$	$\frac{c}{x}(\frac{x}{\tau})^c e^{-(x/\tau)^c}$	$c[(\frac{x}{\tau})^c-1]$	$\frac{\tau^2}{c^2}$	$\frac{\tau^2}{c^2}[(\frac{x}{\tau})^c-1]^2$
gamma	$(0,\infty)$	$\frac{\gamma^\alpha}{x\Gamma(\alpha)}x^\alpha e^{-\gamma x}$	$\gamma x-\alpha$	$\frac{\alpha}{\gamma^2}$	$(x-\alpha/\gamma)^2$
log-logistic	$(0,\infty)$	$\frac{c}{x}\frac{(x/\tau)^c}{[(x/\tau)^c+1]^2}$	$c\frac{(x/\tau)^c-1}{(x/\tau)^c+1}$	$\frac{3\tau^2}{c^2}$	$\frac{9\tau^2}{c^2}\frac{[(x/\tau)^c-1]^2}{[(x/\tau)^c+1]^2}$
Pareto	(a,∞)	ca^c/x^{c+1}	$c-\frac{a(c+1)}{x}$	$\frac{a^2(c+2)}{c^3}$	$\frac{a^2(c+2)^2}{c^2}\left(1-\frac{a(c+1)}{cx}\right)^2$
beta-prime	$(0,\infty)$	$\frac{1}{B(p,q)}\frac{x^{p-1}}{(x+1)^{p+q}}$	$\frac{qx-p}{x+1}$	$\frac{p(p+q+1)}{q^3}$	$\frac{(p+q+1)^2}{q^2}\frac{(x-p/q)^2}{(x+1)^2}$
beta	$(0,1)$	$\frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}$	$(p+q)x-p$	$\frac{pq(p+q+1)}{(p+q)^4}$	$\frac{(p+q+1)^2}{(p+q)^2}(x-\frac{p}{p+q})^2$

Table 3.1: Uncertainty functions of some distributions.

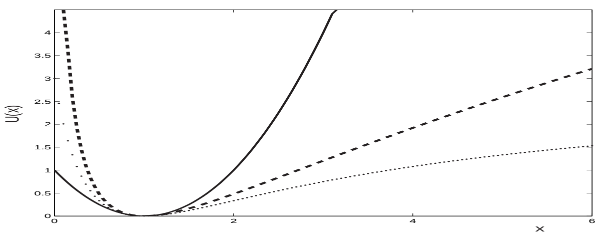

Figure 3.1: Uncertainty functions of gamma (full line), lognormal (dashed line) and log-logistic (dotted line) distributions.

F	$f(x)$	$e^{h(X)}$	$\omega(X)$
normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$\sqrt{2\pi e}\sigma$	σ
Cauchy	$\frac{1}{\pi\sigma(1+(x/\sigma)^2)}$	$4\pi\sigma$	2σ
gamma	$\frac{\gamma^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\gamma x}$	$\frac{\Gamma(\alpha)}{\gamma}e^{((1-\alpha)\psi(\alpha)+\alpha)}$	$\sqrt{\alpha}/\gamma$
Weibull	$\frac{c}{\tau}x^{c-1}e^{-\frac{x^c}{\tau}}$	$\frac{\tau^{1/c}}{c}e^{(c-1)\epsilon/c+1}$	$\frac{\tau^{1/c}}{c}$
Pareto	ca^c/x^{c+1}	$\frac{a}{c}e^{(1+1/c)}$	$\frac{a}{c}\frac{\sqrt{c+2}}{\sqrt{c}}$
power	cx^{c-1}	$\frac{1}{c}e^{(1-1/c)}$	$\frac{\sqrt{c(c+2)}}{(c+1)^2}$

Table 3.2: Comparison of $e^{h(X)}$ and $\omega(X)$ for some distributions.

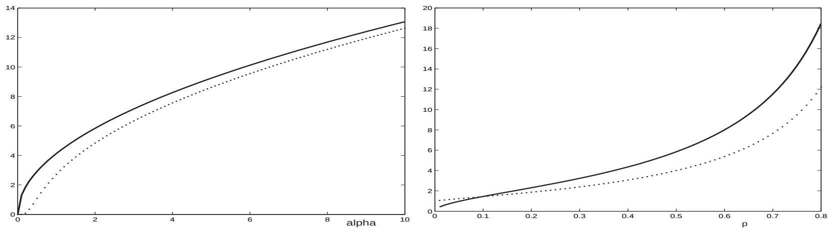

Figure 3.2: $\sqrt{2\pi e}\omega(X)$ (full line) and $e^{h(X)}$ (dotted line) of gamma distribution as function of α (left) and of geometric distribution as function of p (right).

Denote the square root of the mean uncertainty ω^2 by $\omega(X)$. Instead of the differential entropy $h(X)$, the positive values $e^{h(X)}$ are sometimes studied (see [2]). Table 3 shows a close relation between $\omega(X)$ and $e^{h(X)}$ of distributions with support $\mathbb{R}$. The correspondence between $\omega(X)$ and $e^{h(X)}$ of distributions with support $\mathcal{X} \neq \mathbb{R}$ is less apparent, but they have, generally, a similar behavior. As an example, the left panel of Fig. 3 shows $e^{h(X)}$ and $\sqrt{2\pi e}\omega(X)$ as functions of parameter α of the gamma distribution.

4 Uncertainty function of discrete random variables

In the last section we generalize the concept of the t-score for discrete distributions and show that the logarithm of the mean uncertainty has similar behavior as Shannon entropy.

Let a random variable takes on values $k = 0, 1, 2, ...$ with probabilities $f(k)$. As an analogy with distributions with support $\mathcal{X} = (0, \infty)$, for which $1/\eta'(x) = x$, the t-score of the discrete distribution can be determined by replacing in formula (2.4) the derivatives by differences,

$$T(k) = -\frac{1}{f(k)}[(k+1)f(k+1) - kf(k)] = k - (k+1)\frac{f(k+1)}{f(k)}. \tag{4.1}$$

Example 4.1. *Geometric distribution* has probability mass function $f(k) = (1-p)p^k$. By (4.1), $T(k) = k(1-p) - p$. The t-mean $k^* = \frac{p}{1-p}$ equals the mean, and, since $\mathrm{E}T^2 = p$, $\omega^2 = (k^*)^2/\mathrm{E}T^2 = p/(1-p)^2$, which is the ordinary variance. The uncertainty function is, by (3.2),

$$U(k) = \frac{\omega^2}{\mathrm{E}T^2}T^2(k) = \left(k - \frac{p}{1-p}\right)^2.$$

Functions of p, $e^{H(X)}$, where $H(X) = -\log p - \frac{p}{1-p}\log p$ and $\sqrt{2\pi e}\omega(X)$ of the geometric distribution are similar (right panel of Fig. 3).

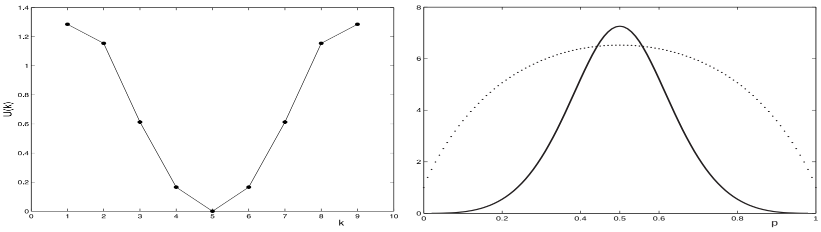

Figure 4.1: Binomial distribution, $n = 10, p = 0.5$. Left: Uncertainty function, right: $\sqrt{2\pi e}\omega(X)$ (full line) and $e^{H(X)}$ (dotted line).

Example 4.2. *Poisson distribution* has probability mass function $f(k) = \frac{e^{-\lambda}\lambda^k}{k!}$. By (4.1), $T(k) = k - \lambda$, $x^* = \omega^2 = \lambda$ and

$$U(k) = (k - \lambda)^2.$$

Let n be a fixed number and random variable X takes on values $k = 0, 1, 2, ..., n$ with probabilities $f(k)$. As an analogy with distributions with finite interval support $\mathcal{X} = (0, n)$, for which $1/\eta'(x) = x(n-x)/n$, the t-score of the discrete distribution can be written as

$$T(x) = \frac{1}{f(x)}\frac{d}{dx}\left[-\frac{x(n-x)}{n}f(x)\right]. \tag{4.2}$$

If we approximate (4.2) by symmetric differences, we obtain

$$T(k) = -1 + \frac{2k}{n} - \frac{k(n-k)}{2nf(k)}[f(k+1) - f(k-1)] \tag{4.3}$$

for $k = 1, ..., n-1$, with $T(0) = -1$ and $T(n) = 1$. The score variance is then

$$\omega^2 = \frac{[k^*(n-k^*)]^2}{n^2 \mathrm{E}T^2}. \tag{4.4}$$

Example 4.3. *Discrete uniform distribution* has probabilities $f(k) = \frac{1}{n+1}$. Its t-score is $T(k) = 2k/n - 1$ so that $x^* = n/2$ equals the mean. The t-score moment is $\mathrm{E}T^2 = 2(2n+1)/3n - 1$. For large n, $\mathrm{E}T^2 \doteq 1/3$. The score variance is $\omega^2 = \frac{n^2}{2^4 \mathrm{E}T^2}$ and the uncertainty function

$$U(k) = \frac{(k - n/2)^2}{4[(2n+1)/3n - 1)]^2}.$$

For large n, $U(k) \doteq \frac{9}{4}(k - n/2)^2$. The uncertainty function of the continuous uniform distribution on $0, 1)$ is $U(x) = \frac{9}{4}(x - 1/2)^2$.

Example 4.4. *Binomial distribution* has mass probability function $f(k) = \binom{n}{k}p^k(1-p)^{n-k}$. By (4.3), its t-score is

$$T(k) = -1 + \frac{2k}{n} - \frac{k(n-k)}{2n}\left[\frac{(n-k)p}{(k+1)(1-p)} - \frac{k(1-p)}{(n-k+1)p}\right].$$

Using (3.2), (4.4), numerical solutions of equation $T(x) = 0$ and the direct computation of $\frac{1}{n+1}\sum_{k=0}^{n} T^2(k)$, we obtained a plot of uncertainty function for $n = 10$ and $p = 0.5$ (left panel of Fig. 4), and the comparison of functions of p, the square root of the mean uncertainty $\omega(X)$ multiplied by term $\sqrt{2\pi e}$ with $e^{H(X)}$ (right panel of Fig.4). In this case, the mean uncertainty seems to be a better tool for distinguishing values of p as the Shannon entropy.

Aknowledgements. The research presented in the paper was supported by grant ME 949 "Analysis of negative impacts on driver attention" of The Ministry of Education, Youth and Sports of the Czech Republic.

Bibliography

[1] Balakrishnan, N., and Nevzorov, V. B., *A Primer on Statistical Distributions*, Wiley, Hoboken (2003).

[2] Cover, T. M., and Thomas, J.A., *Elements of Information Theory*, Wiley, Hoboken (1991).

[3] Fabián, Z., "Information and entropy of continuous random variables", *IEEE Trans. of Information Theory* 43, 1080–1084, (1997).

[4] Fabián, Z., "Induced cores and their use in robust parametric estimation", *Communication in Statistics, Theory Methods 30, 537-556 (2001).*

[5] Fabián, Z., "New measures of central tendency and variability of continuous distributions". Communication in Statistics, Theory Methods *37, 159–174 (2008).*

[6] Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. Robust Statistic. The Approach Based on Influence Functions, *Wiley, New York (1986).*

[7] Johnson, N. L., "Systems of frequency curves generated by methods of translations". Biometrika *36: 149–176 (1949).*

[8] Lawless J. F., Statistical models and methods for lifetime data*, Wiley, New York (2003).*