p. 14

TDA Progress Report 42-104 February 15, 1991



# Maximum-Entropy Probability Distributions Under $L_p$-Norm Constraints

S. Dolinar
Communications Systems Research Section

*This article tabulates continuous probability density functions and discrete probability mass functions which maximize the differential entropy or absolute entropy, respectively, among all probability distributions with a given $L_p$-norm (i.e., a given $p$th absolute moment when $p$ is a finite integer) and unconstrained or constrained value set. Expressions for the maximum entropy are evaluated as functions of the $L_p$-norm. The most interesting results are obtained and plotted for unconstrained (real-valued) continuous random variables and for integer-valued discrete random variables.*

*The maximum entropy expressions are obtained in closed form for unconstrained continuous random variables, and in this case there is a simple straight-line relationship between the maximum differential entropy and the logarithm of the $L_p$-norm. Corresponding expressions for arbitrary discrete and constrained continuous random variables are given parametrically; closed-form expressions are available only for special cases. However, simpler alternative bounds on the maximum entropy of integer-valued discrete random variables are obtained by applying the differential entropy results to continuous random variables which approximate the integer-valued random variables in a natural manner.*

*Most of these results are not new. The purpose of this article is to present all the results in an integrated framework that includes continuous and discrete random variables, constraints on the permissible value set, and all possible values of p. Understanding such as this is useful in evaluating the performance of data compression schemes.*

## I. Introduction

The *differential entropy* $\mathbf{h}\{x\}$ of a continuous, real-valued random variable $x$ with probability density $f(x)$ is defined as

$$\mathbf{h}\{x\} = -\mathbf{E}\{\log[f(x)]\} = -\int_{-\infty}^{\infty} f(x)\log[f(x)]dx \quad (1)$$

For any positive (or infinite) integer $p = 1, 2, 3, \ldots, \infty$, define the $L_p$-norm $\mathbf{M}_p\{x\}$ of the random variable $x$ as

$$\begin{aligned}\mathbf{M}_p\{x\} &= [\mathbf{E}\{|x|^p\}]^{1/p} \\ &= \left[\int_{-\infty}^{\infty} f(x)|x|^p dx\right]^{1/p}, \quad p = 1, 2, 3, \ldots \\ \mathbf{M}_\infty\{x\} &= \lim_{p\to 0} \mathbf{M}_p\{x\} = \operatorname*{ess\,sup}_{f(x)>0}|x| \end{aligned} \quad (2)$$

The essential supremum in Eq. (2) is the smallest number that upper bounds $|x|$ almost surely.

Sometimes the real-valued random variable $x$ is constrained to lie within a subset $\Xi$ of the real line; in this case, the integrals in Eqs. (1) and (2) need only extend over the subset $\Xi$.

For a discrete random variable $X$ with discrete value set $\Xi = \{\xi_i\}$ and probability mass function $F(\xi_i)$, its (absolute) *entropy* $\mathbf{H}\{X\}$ is defined as

$$\mathbf{H}\{X\} = -\mathbf{E}\{\log[F(X)]\} = -\sum_i F(\xi_i)\log[F(\xi_i)] \quad (3)$$

The $L_p$-norm $\mathbf{M}_p\{X\}$ of the discrete random variable $X$ is defined as

$$\begin{aligned}\mathbf{M}_p\{X\} &= [\mathbf{E}\{|X|^p\}]^{1/p} \\ &= \left[\sum_i F(\xi_i)|\xi_i|^p\right]^{1/p}, \quad p = 1, 2, 3, \ldots \\ \mathbf{M}_\infty\{X\} &= \lim_{p\to\infty} \mathbf{M}_p\{X\} = \sup_{F(\xi_i)>0} |\xi_i| \end{aligned} \quad (4)$$

This article tabulates continuous probability density functions $f(x) = f_p^*(x;\mu)$ or $f(x) = f_p^*(x;\mu,\Xi)$ and discrete probability mass functions $F(\xi_i) = F_p^*(\xi_i;\mu,\Xi)$ which maximize the differential entropy $\mathbf{h}\{x\}$ or absolute entropy $\mathbf{H}\{X\}$, respectively, among all probability distributions with a given $L_p$-norm $\mathbf{M}_p\{x\}$ or $\mathbf{M}_p\{X\}$ and unconstrained or constrained value set $\Xi$. The most interesting results are obtained and plotted for unconstrained continuous random variables and for integer-valued discrete random variables. Finally, alternative simpler bounds on the entropy of integer-valued random variables are obtained by modifying the bounds on differential entropy for unconstrained continuous random variables.

Most of these results are not new. In fact, the maximum-entropy continuous distributions for $p = 1, 2$ (Laplacian and Gaussian distributions, respectively) have been known since Shannon's original work [1]. The purpose of this article is to present all the results in an integrated framework that includes continuous and discrete random variables, constraints on the permissible value set, and all possible values of $p$.

Throughout this article, regular italic notation is used for an ordinary function of a real variable, such as $f(x)$ or $F(\xi_i)$, while boldface notation is used for an operator applied to a random variable, such as $\mathbf{h}\{x\}$ or $\mathbf{H}\{X\}$, $\mathbf{M}_p\{x\}$ or $\mathbf{M}_p\{X\}$, or the expectation operator $\mathbf{E}\{\cdot\}$. In order not to interrupt the main presentation, proofs of all stated results are relegated to the Appendix.

## II. Effects of Elementary Transformations

A *scaled* random variable $x' = qx$ or $X' = qX$, where $q$ is a constant, has a correspondingly scaled $L_p$-norm:

$$\begin{aligned}\mathbf{M}_p\{x'\} &= |q|\mathbf{M}_p\{x\} \\ \mathbf{M}_p\{X'\} &= |q|\mathbf{M}_p\{X\}\end{aligned} \quad (5)$$

A discrete random variable $X$ with value set $\Xi = \{\xi_i\}$ scales to a discrete random variable $X'$ with scaled value set $q\Xi \equiv \{q\xi_i\}$. The entropy of a discrete random variable is unaffected by scaling, but the differential entropy of a scaled continuous random variable either increases or decreases:

$$\begin{aligned}\mathbf{h}\{x'\} &= \mathbf{h}\{x\} + \log[|q|] \\ \mathbf{H}\{X'\} &= \mathbf{H}\{X\}\end{aligned} \quad (6)$$

The change in the differential entropy of a scaled continuous random variable exactly equals the change in the logarithm of its $L_p$-norm:

$$\mathbf{h}\{x'\} - \mathbf{h}\{x\} = \log[\mathbf{M}_p\{x'\}] - \log[\mathbf{M}_p\{x\}] = \log[|q|] \quad (7)$$

In contrast, the $L_p$-norm of a discrete random variable can be made arbitrarily small or large without affecting its entropy, simply by scaling its value set.

A *shifted* random variable $x'' = x - \Delta$ or $X'' = X - \Delta$, where $\Delta$ is a constant, has the same differential or absolute entropy as the unshifted random variable,

$$\mathbf{h}\{x''\} = \mathbf{h}\{x\}$$
$$\mathbf{H}\{X''\} = \mathbf{H}\{X\} \tag{8}$$

but a different $L_p$-norm. A discrete random variable $X$ with value set $\Xi = \{\xi_i\}$ shifts to a discrete random variable $X''$ with shifted value set $\Xi - \Delta \equiv \{\xi_i - \Delta\}$. A random variable $x$ or $X$ is *centered* with respect to the $L_p$-norm if no shifted version has a lower $L_p$-norm. A centered random variable $x_p^o$ or $X_p^o$ can be obtained from an uncentered random variable $x$ or $X$ by applying an optimum shift $\Delta = \Delta_p^o$. This optimum shift equals the median of the random variable for $p = 1$, the mean value of the random variable for $p = 2$, and the average of the essential infimum and essential supremum of the random variable for $p = \infty$. The centered $L_p$-norm $\mathbf{M}_p^o\{x\}$ or $\mathbf{M}_p^o\{X\}$ of the random variable $x$ or $X$ can be defined as

$$\mathbf{M}_p^o\{x\} = \min_{\Delta} \mathbf{M}_p\{x - \Delta\} = \mathbf{M}_p\left\{x - \Delta_p^o\right\} = \mathbf{M}_p\{x_p^o\}$$
$$\mathbf{M}_p^o\{X\} = \min_{\Delta} \mathbf{M}_p\{X - \Delta\} = \mathbf{M}_p\left\{X - \Delta_p^o\right\} = \mathbf{M}_p\{X_p^o\} \tag{9}$$

## III. Maximum Differential Entropy for Continuous Random Variables

For any positive real number $\mu$ and any positive (or infinite) integer $p = 1, 2, \ldots, \infty$, let $x_p^*(\mu)$ be a continuous random variable with probability density $f_p^*(x;\mu)$, where

$$f_p^*(x;\mu) = \frac{\exp(-|x|^p/p\mu^p)}{2\mu p^{1/p}\,\Gamma\left(\frac{p+1}{p}\right)}, \quad p = 1, 2, 3, \ldots$$

$$f_\infty^*(x;\mu) = \begin{cases} \dfrac{1}{2\mu}, & |x| \le \mu \\ 0, & |x| > \mu \end{cases} \tag{10}$$

and $\Gamma(\cdot)$ is the gamma function. These probability densities are all properly normalized, i.e.,

$$\int_{-\infty}^{\infty} f_p^*(x;\mu)\,dx = 1, \ p = 1, 2, 3, \ldots, \infty \tag{11}$$

The probability densities $f_p^*(x;\mu)$ for $p = 1, 2, \infty$ are the well-known Laplacian, Gaussian, and uniform probability densities, respectively.

The absolute moments of these random variables are known in closed form:

$$\mathbf{E}\{|x_p^*(\mu)|^n\} = \mu^n \frac{\Gamma\left(\frac{n+1}{p}\right) p^{n/p}}{\Gamma\left(\frac{1}{p}\right)},$$
$$n = 1, 2, 3, \ldots, \quad p = 1, 2, 3, \ldots$$
$$\mathbf{E}\{|x_\infty^*(\mu)|^n\} = \frac{\mu^n}{n+1}, \quad n = 1, 2, 3, \ldots \tag{12}$$

Evaluating these expressions for $n = p$ or $n \to \infty$ yields the $L_p$-norm $M_p^*(\mu)$ of the random variable $x_p^*(\mu)$:

$$M_p^*(\mu) \equiv \mathbf{M}_p\{x_p^*(\mu)\} = \mu, \quad p = 1, 2, 3, \ldots, \infty \tag{13}$$

The differential entropy $h_p^*(\mu)$ of the random variable $x_p^*(\mu)$ is calculated as

$$h_p^*(\mu) \equiv \mathbf{h}\{x_p^*(\mu)\}$$
$$= \log\left[2\mu\,\Gamma\left(\tfrac{p+1}{p}\right)(pe)^{1/p}\right], \quad p = 1, 2, 3, \ldots$$
$$h_\infty^*(\mu) \equiv \mathbf{h}\{x_\infty^*(\mu)\} = \log[2\mu] \tag{14}$$

Explicit formulas for $p = 1, 2$ are

$$h_1^*(\mu) = \log[2e\mu]$$
$$h_2^*(\mu) = \log[\sqrt{2\pi e}\;\mu] \tag{15}$$

Since from Eq. (13) the parameter $\mu$ equals the $L_p$-norm $M_p^*(\mu)$ for any $p$, the differential entropy can be related directly to the corresponding $L_p$-norm:

$$h_p^*(\mu) = \log\left[2\,\Gamma\left(\tfrac{p+1}{p}\right)(pe)^{1/p}\right] + \log[M_p^*(\mu)],$$
$$p = 1, 2, 3, \ldots$$
$$h_\infty^*(\mu) = \log[2] + \log\left[M_\infty^*(\mu)\right] \tag{16}$$

The differential entropy $h_p^*(\mu)$ is plotted in Fig. 1 versus the logarithm of the corresponding $L_p$-norm, $\log[M_p^*(\mu)]$, for various values of $p$. Note that this is a simple straight-line relationship. In fact, the straight line has unit slope, assuming $\log[M_p^*(\mu)]$ is measured to the same logarithmic base as $h_p^*(\mu)$. This is consistent with the previous observation in Eq. (7), because the scaled version of the random variable $x_p^*(\mu)$ is statistically equivalent to the random variable with scaled $L_p$-norm, i.e.,

$$q x_p^*(\mu) \Leftrightarrow x_p^*(|q|\mu) \tag{17}$$

If $x$ is any continuous random variable with differential entropy $\mathbf{h}\{x\}$ and $L_p$-norm $\mathbf{M}_p\{x\} = \mu$, then

$$\mathbf{h}\{x\} \le h_p^*(\mathbf{M}_p\{x\}) = \mathbf{h}\{x_p^*(\mu)\}, \ p = 1, 2, 3, \ldots, \infty \tag{18}$$

i.e., $x_p^*(\mu)$ is the *maximum-entropy* continuous random variable with a fixed $L_p$-norm $\mu$. Since the bound in Eq. (19) must be valid for all values of $p$,

$$\mathbf{h}\{x\} \le \min_p h_p^*(\mathbf{M}_p\{x\}) \tag{19}$$

If the random variable $x$ is not centered with respect to the $L_p$-norm, the centered random variable $x_p^o = x - \Delta_p^o$ has the same differential entropy as $x$ but a smaller $L_p$-norm. The differential entropy of $x$ may be more tightly upper bounded by applying the bounds in Eqs. (18) and (19) to the differential entropy of $x_p^o$:

$$\mathbf{h}\{x\} = \mathbf{h}\{x_p^o\} \le h_p^*(\mathbf{M}_p\{x_p^o\}) = h_p^*(\mathbf{M}_p^o\{x\}),$$
$$p = 1, 2, 3, \ldots, \infty \tag{20}$$

and

$$\mathbf{h}\{x\} \le \min_p h_p^*(\mathbf{M}_p^o\{x\}) \tag{21}$$

If the real-valued continuous random variable $x$ is constrained to lie within a subset $\Xi$ of the real line, its maximum possible differential entropy is smaller than that calculated above for a random variable constrained only by its $L_p$-norm. Maximum-entropy distributions for constrained continuous random variables can be obtained as simple generalizations of the foregoing results. Let $x_p^*(\mu, \Xi)$ be a continuous random variable with probability density $f_p^*(x; \mu, \Xi)$ equal to the conditional probability density of $x_p^*(\mu)$ given $\{x_p^*(\mu) \in \Xi\}$, i.e.,

$$f_p^*(x;\mu,\Xi) = \begin{cases} \dfrac{\exp(-|x|^p/p\mu^p)}{\alpha_p^*(\mu,\Xi)}, & x \in \Xi \\ & \qquad p = 1,2,3, \ldots \\ 0, & x \notin \Xi \end{cases}$$

$$f_\infty^*(x;\mu,\Xi) = \begin{cases} \dfrac{1}{\alpha_\infty^*(\mu,\Xi)}, & |x| \le \mu \text{ and } x \in \Xi \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

where

$$\alpha_p^*(\mu,\Xi) = \int_\Xi \exp(-|x|^p/p\mu^p)\ dx, \ p = 1, 2, 3, \ldots$$

$$\alpha_\infty^*(\mu,\Xi) = \int_{\Xi\cap\{|x|\le\mu\}} 1\ dx \tag{23}$$

The $L_p$-norm $M_p^*(\mu,\Xi)$ of the random variable $x_p^*(\mu,\Xi)$ is given by

$$M_p^*(\mu,\Xi) \equiv \mathbf{M}_p\{x_p^*(\mu,\Xi)\}$$
$$= \mu\left[\frac{\beta_p^*(\mu,\Xi)}{\alpha_p^*(\mu,\Xi)}\right]^{\frac{1}{p}}, \ p = 1, 2, 3, \ldots$$
$$M_\infty^*(\mu,\Xi) \equiv \mathbf{M}_\infty\{x_\infty^*(\mu,\Xi)\} = \sup_{\substack{|x|\le\mu \\ x\in\Xi}} |x| \tag{24}$$

where

$$\beta_p^*(\mu,\Xi) = \int_\Xi (|x|^p/\mu^p)\exp(-|x|^p/p\mu^p)\ dx,$$
$$p = 1, 2, 3, \ldots \tag{25}$$

The differential entropy $h_p^*(\mu,\Xi)$ of the random variable $x_p^*(\mu,\Xi)$ is given by

$$h_p^*(\mu,\Xi) \equiv \mathbf{h}\{x_p^*(\mu,\Xi)\}$$
$$= \log[\alpha_p^*(\mu,\Xi)] + \frac{\log[e]}{p}\frac{\beta_p^*(\mu,\Xi)}{\alpha_p^*(\mu,\Xi)}$$
$$= \log[\alpha_p^*(\mu,\Xi)] + \frac{\log[e]}{p}\left[\frac{M_p^*(\mu,\Xi)}{\mu}\right]^p,$$
$$p = 1, 2, 3, \ldots$$

$$h_\infty^*(\mu,\Xi) \equiv \mathbf{h}\{x_\infty^*(\mu,\Xi)\} = \log[\alpha_\infty^*(\mu,\Xi)] \tag{26}$$

The random variable $x_p^*(\mu,\Xi)$ is the *maximum-entropy* continuous random variable with constrained value set $X$ and fixed $L_p$-norm $M_p^*(\mu,\Xi)$, i.e., if $x$ is any continuous random variable with value set $\Xi$, differential entropy $\mathbf{h}\{x\}$, and $L_p$-norm $\mathbf{M}_p\{x\}$, then

$$\mathbf{h}\{x\} \le \mathbf{h}\{x_p^*(\mu_p,\Xi)\}$$
$$= h_p^*(\mu_p,\Xi), \ p = 1, 2, 3, \ldots, \infty \tag{27}$$

where $\mu_p$ is chosen to match the $L_p$-norm of $x$:

$$M_p^*(\mu_p, \Xi) = \mathbf{M}_p\{x\}, \; p = 1, 2, 3, \ldots, \infty \tag{28}$$

Since the bound in Eq. (27) must be valid for all values of $p$,

$$\mathbf{h}\{x\} \leq \min_p h_p^*(\mu_p, \Xi) \tag{29}$$

If the random variable $x$ is not centered with respect to the $L_p$-norm, the differential entropy of $x$ may be more tightly upper bounded by applying the bounds in Eqs. (27) and (29) to the differential entropy of the centered random variable $x_p^o = x - \Delta_p^o$:

$$\mathbf{h}\{x\} = \mathbf{h}\{x_p^o\} \leq \mathbf{h}\{x_p^*(\mu_p^o, \Xi - \Delta_p^o)\} = h_p^*(\mu_p^o, \Xi - \Delta_p^o),$$

$$p = 1, 2, 3, \ldots, \infty \tag{30}$$

and

$$\mathbf{h}\{x\} \leq \min_p h_p^*(\mu_p^o, \Xi - \Delta_p^o) \tag{31}$$

where $\mu_p^o$ is chosen to match the $L_p$-norm of $x_p^o$ (i.e., the centered $L_p$-norm of $x$):

$$M_p^*(\mu_p^o, \Xi - \Delta_p^o) = \mathbf{M}_p\{x_p^o\}$$

$$= \mathbf{M}_p^o\{x\}, \; p = 1, 2, 3, \ldots, \infty \tag{32}$$

Notice that the bounds on the right-hand sides of Eqs. (30) and (31) are calculated with reference to the shifted value sets $\Xi - \Delta_p^o$, not the actual value set $\Xi$.

The integrals defining $\alpha_p^*(\mu, \Xi)$ and $\beta_p^*(\mu, \Xi)$ are generally not obtainable in closed form for an arbitrary value set $\Xi$. An interesting exception is when the value set equals the positive half-line, i.e., $\Xi = R^+ \equiv (0, \infty)$. In this case,

$$M_p^*(\mu, R^+) = M_p^*(\mu) = \mu, \; p = 1, 2, 3, \ldots, \infty \tag{33}$$

and

$$h_p^*(\mu, R^+) = h_p^*(\mu) - \log[2], \; p = 1, 2, 3, \ldots, \infty \tag{34}$$

In other words, the maximum possible differential entropy for a positive-valued continuous random variable is exactly one bit less than the maximum differential entropy for a real-valued random variable with the same $L_p$-norm.

## IV. Maximum Entropy for Discrete Random Variables

Discrete versions $F_p^*(\xi_i; \mu, \Xi)$ of the probability densities $f_p^*(x; \mu)$ can be defined in a natural manner for discrete random variables $X_p^*(\mu, \Xi)$ with discrete value set $\Xi = \{\xi_i\}$:

$$F_p^*(\xi_i; \mu, \Xi) = \frac{f_p^*(\xi_i; \mu)}{\sum_j f_p^*(\xi_j; \mu)}$$

$$= \frac{\exp(-|\xi_i|^p / p\mu^p)}{A_p^*(\mu, \Xi)}, \; p = 1, 2, 3, \ldots \tag{35}$$

$$F_\infty^*(\xi_i; \mu, \Xi) = \frac{f_\infty^*(\xi_i; \mu)}{\sum_j f_\infty^*(\xi_j; \mu)} = \begin{cases} \dfrac{1}{A_\infty^*(\mu, \Xi)}, & |\xi_i| \leq \mu \\ 0, & |\xi_i| > \mu \end{cases}$$

where

$$A_p^*(\mu, \Xi) = \sum_i \exp(-|\xi_i|^p / p\mu^p), \; p = 1, 2, 3, \ldots$$

$$A_\infty^*(\mu, \Xi) = \sum_{|\xi_i| \leq \mu} 1 \tag{36}$$

The discrete probability mass function $F_p^*(\xi_i; \mu, \Xi)$ equals the conditional probability mass function for the maximum-entropy continuous random variable $x_p^*(\mu)$, given $\{x_p^*(\mu) \in \Xi\}$.

The $L_p$-norm $M_p^*(\mu, \Xi)$ of the discrete random variable $X_p^*(\mu, \Xi)$ is given by

$$M_p^*(\mu, \Xi) \equiv \mathbf{M}_p\{X_p^*(\mu, \Xi)\}$$

$$= \mu \left[ \frac{B_p^*(\mu, \Xi)}{A_p^*(\mu, \Xi)} \right]^{\frac{1}{p}}, \; p = 1, 2, 3, \ldots$$

$$M_\infty^*(\mu, \Xi) \equiv \mathbf{M}_\infty\{X_\infty^*(\mu, \Xi)\} = \sup_{|\xi_i| \leq \mu} |\xi_i| \tag{37}$$

where

$$B_p^*(\mu, \Xi) = \sum_i (|\xi_i|^p / \mu^p) \exp(-|\xi_i|^p / p\mu^p),$$

$$p = 1, 2, 3, \ldots \tag{38}$$

The entropy $H_p^*(\mu,\Xi)$ of the discrete random variable $X_p^*(\mu,\Xi)$ is given by

$$H_p^*(\mu,\Xi) \equiv \mathbf{H}\{X_p^*(\mu,\Xi)\}$$

$$= \log[A_p^*(\mu,\Xi)] + \frac{\log[e]}{p} \frac{B_p^*(\mu,\Xi)}{A_p^*(\mu,\Xi)}$$

$$= \log[A_p^*(\mu,\Xi)] + \frac{\log[e]}{p} \left[\frac{M_p^*(\mu,\Xi)}{\mu}\right]^p,$$

$$p = 1,2,3,\ldots$$

$$H_\infty^*(\mu,\Xi) \equiv \mathbf{H}\{X_\infty^*(\mu,\Xi)\} = \log[A_\infty^*(\mu,\Xi)] \qquad (39)$$

The random variable $X_p^*(\mu,\Xi)$ is the *maximum-entropy* discrete random variable with value set $\Xi$ and fixed $L_p$-norm $M_p^*(\mu,\Xi)$; i.e., if $X$ is any discrete random variable with value set $\Xi$, entropy $\mathbf{H}\{X\}$, and $L_p$-norm $\mathbf{M}_p\{X\}$, then

$$\mathbf{H}\{X\} \leq \mathbf{H}\{X_p^*(\mu_p,\Xi)\}$$

$$= H_p^*(\mu_p,\Xi),\ p = 1,2,3,\ldots,\infty \qquad (40)$$

where $\mu_p$ is chosen to match the $L_p$-norm of $X$:

$$M_p^*(\mu_p,\Xi) = \mathbf{M}_p\{X\},\ p = 1,2,3,\ldots,\infty \qquad (41)$$

Since the bound in Eq. (40) must be valid for all values of $p$,

$$\mathbf{H}\{X\} \leq \min_p H_p^*(\mu_p,\Xi) \qquad (42)$$

If the random variable $X$ is not centered with respect to the $L_p$-norm, the centered random variable $X_p^o = X - \Delta_p^o$ has the same entropy as $X$ but a smaller $L_p$-norm. The entropy of $X$ may be more tightly upper bounded by applying the bounds in Eqs. (40) and (42) to the entropy of $X_p^o$.

$$\mathbf{H}\{X\} = \mathbf{H}\{X_p^o\}$$

$$\leq \mathbf{H}\{X_p^*(\mu_p^o,\Xi - \Delta_p^o)\}$$

$$= H_p^*(\mu_p^o,\Xi - \Delta_p^o),\ p = 1,2,3,\ldots,\infty \qquad (43)$$

and

$$\mathbf{H}\{X\} \leq \min_p H_p^*(\mu_p^o,\Xi - \Delta_p^o) \qquad (44)$$

where $\mu_p^o$ is chosen to match the $L_p$-norm of $X_p^o$ (i.e., the centered $L_p$-norm of $X$):

$$M_p^*(\mu_p^o,\Xi - \Delta_p^o) = \mathbf{M}_p\{X_p^o\}$$

$$= \mathbf{M}_p^o\{X\},\ p = 1,2,3,\ldots,\infty \qquad (45)$$

Notice again that the bounds based on centered random variables are calculated with reference to the shifted value sets $\Xi - \Delta_p^o$, not the actual value set $\Xi$. An exception for which the centering operation leaves the value set unchanged (i.e., $\Xi - \Delta_p^o = \Xi$) occurs for the value set $\Xi = I$ (defined below) or, more generally, for any scaled version of it, $\Xi = qI$ , as long as the allowable centering shifts $\Delta_p^o$ are constrained to multiples of the scale quantum $q$.

For many applications, the most interesting discrete value sets are the set of all integers $I \equiv \{0,\pm1,\pm2,\pm3,\ldots\}$ and the set of positive integers $I^+ \equiv \{1,2,3,\ldots\}$. The maximum entropy for integer-valued random variables, $H_p^*(\mu,I)$, is plotted in Fig. 2 versus the logarithm of the corresponding $L_p$-norm, $\log[M_p^*(\mu,I)]$, for various values of $p$. Notice that the nonlinear relationship for integer-valued random variables becomes essentially linear when the $L_p$-norm is large compared to the (unit) interval between successive values in the value set $I$. In fact, all of the curves in Fig. 2 converge to the corresponding straight-line curves in Fig. 1 in the limit of large $L_p$-norm. Notice also how the continuous curves for large values of $p < \infty$ approach the limiting staircase curve for $p = \infty$. The maximum entropy curve for $p = \infty$ takes quantum jumps at integer values of the $L_\infty$-norm.

Closed-form maximum-entropy expressions as a function of $L_p$-norm can be obtained for discrete random variables in only a few special cases. Interesting cases include $p = 1,\infty$, for value sets $\Xi = I, I^+$:

$$H_1^*(\mu,I) = \log\left[M_1^*(\mu,I) + \sqrt{1 + [M_1^*(\mu,I)]^2}\right]$$

$$+ M_1^*(\mu,I)\log\left[\frac{M_1^*(\mu,I)}{\sqrt{1 + [M_1^*(\mu,I)]^2} - 1}\right]$$

$$= M_1^*(\mu, I)$$

$$\times \mathsf{H}_2 \left[ \frac{1 + M_1^*(\mu, I) - \sqrt{1 + [M_1^*(\mu, I)]^2}}{2M_1^*(\mu, I)} \right]$$

$$+ \frac{1 + M_1^*(\mu, I) + \sqrt{1 + [M_1^*(\mu, I)]^2}}{2}$$

$$\times \mathsf{H}_2 \left[ \sqrt{1 + [M_1^*(\mu, I)]^2} - M_1^*(\mu, I) \right]$$

$$H_\infty^*(\mu, I) = \log(2\lfloor M_\infty^*(\mu, I) \rfloor + 1) \tag{46}$$

and

$$H_1^*(\mu, I^+) = M_1^*(\mu, I^+) \mathsf{H}_2 \left[ \frac{1}{M_1^*(\mu, I^+)} \right]$$

$$H_\infty^*(\mu, I^+) = \log\left(\lfloor M_\infty^*(\mu, I^+) \rfloor\right) \tag{47}$$

where $\lfloor a \rfloor$ is the integer part of $a$ and $\mathsf{H}_2[a]$ is the binary entropy function,

$$\mathsf{H}_2[a] = \begin{cases} -a \log[a] - (1-a)\log[1-a], & 0 < a < 1 \\ 0, & a = 0 \text{ or } a = 1 \end{cases} \tag{48}$$

## V. Alternative Entropy Bounds for Integer-Valued Random Variables

The maximum-entropy discrete distributions are not as useful as the maximum-entropy continuous distributions for unconstrained value sets, because closed-form results determining the maximum entropy for a given $L_p$-norm are available only in special cases. Alternative bounds on the entropy of discrete random variables can be obtained by approximating their discrete probability distributions with continuous probability densities and applying the simpler bounds on the differential entropy of continuous random variables. In this section, entropy bounds of this kind are obtained for integer-valued random variables ($\Xi = I$).

Associate with any integer-valued random variable $X$ a corresponding continuous random variable $x$ defined by

$$x = X + u \tag{49}$$

where $u$ is a uniform (continuous) random variable over $[-1/2, 1/2]$ which is independent of $X$. The probability density function $f(x)$ of the continuous random variable $x$ is related to the probability mass function $F(X)$ of the discrete random variable $X$ as:

$$f(x) = F(\lfloor x + 1/2 \rfloor) \tag{50}$$

where $\lfloor x + 1/2 \rfloor$ maps $x$ to the nearest integer. The differential entropy of $x$ equals the absolute entropy of $X$, i.e.,

$$\mathbf{h}\{x\} = \mathbf{H}\{X\} \tag{51}$$

and their $L_p$-norms are related as follows:

$$[\mathbf{M}_p\{x\}]^p = \begin{cases} \displaystyle\sum_{\substack{r=0 \\ r \text{ even}}}^{p-1} [\mathbf{M}_{p-r}\{X\}]^{p-r} \binom{p}{r} \frac{2^{-r}}{r+1} + \frac{2^{-p}}{p+1} F(0), & p = 1, 3, 5, \ldots \\ \displaystyle\sum_{\substack{r=0 \\ r \text{ even}}}^{p-2} [\mathbf{M}_{p-r}\{X\}]^{p-r} \binom{p}{r} \frac{2^{-r}}{r+1} + \frac{2^{-p}}{p+1}, & p = 2, 4, 6, \ldots \end{cases}$$

$$\mathbf{M}_\infty\{x\} = \mathbf{M}_\infty\{X\} + \frac{1}{2} \tag{52}$$

Explicit formulas for $p = 1, 2, 3, 4$, are:

$$\mathbf{M}_1\{x\} = \mathbf{M}_1\{X\} + \frac{1}{4}F(0)$$

$$[\mathbf{M}_2\{x\}]^2 = [\mathbf{M}_2\{X\}]^2 + \frac{1}{12}$$

$$[\mathbf{M}_3\{x\}]^3 = [\mathbf{M}_3\{X\}]^3 + \frac{1}{4}\mathbf{M}_1\{X\} + \frac{1}{32}F(0)$$

$$[\mathbf{M}_4\{x\}]^4 = [\mathbf{M}_4\{X\}]^4 + \frac{1}{2}[\mathbf{M}_2\{X\}]^2 + \frac{1}{80} \tag{53}$$

The entropy of the integer-valued random variable $X$ is upper bounded by

$$\mathbf{H}\{X\} = \mathrm{h}\{x\} \leq h_p^*(\mathbf{M}_p\{x\}), \;\; p = 1, 2, 3, \ldots, \infty \tag{54}$$

Explicit bounds for $p = 1, 2, \infty$, are:

$$\mathbf{H}\{X\} \leq \; \log[2e] + \log\left[\mathbf{M}_1\{X\} + \frac{1}{4}F(0)\right]$$

$$\mathbf{H}\{X\} \leq \; \log[\sqrt{2\pi e}] + \frac{1}{2}\log\left[[\mathbf{M}_2\{X\}]^2 + \frac{1}{12}\right]$$

$$\mathbf{H}\{X\} \leq \log[2] + \log\left[\mathbf{M}_\infty\{X\} + \frac{1}{2}\right] \tag{55}$$

Since the bound in Eq. (54) is valid for all values of $p$,

$$\mathbf{H}\{X\} \leq \min_p h_p^*(\mathbf{M}_p\{x\}) \tag{56}$$

The bound in Eq. (54) is not quite as tight as the achievable bound given earlier in Eq. (40), because the stepwise constant probability density of $x = X + u$ given by Eq. (50) cannot exactly equal the maximum-entropy continuous probability density specified by Eq. (10). However, a stepwise-constant approximation can be very accurate when the probability distribution is much wider than the unit step width.

## VI. Summary and Potential Applications

This article has tabulated continuous probability density functions $f(x) = f_p^*(x;\mu)$ or $f(x) = f_p^*(x;\mu,\Xi)$ and discrete probability mass functions $F(\xi_i) = F_p^*(\xi_i;\mu,\Xi)$ which maximize the differential entropy $\mathrm{h}\{x\}$ or absolute entropy $\mathbf{H}\{X\}$, respectively, among all probability distributions with a given $L_p$-norm $\mathbf{M}_p\{x\}$ or $\mathbf{M}_p\{X\}$ and unconstrained or constrained value set $\Xi$. Expressions for the maximum entropy are evaluated as functions of the $L_p$-norm. These expressions are obtained in closed form for the case of unconstrained continuous random variables, and in this case there is a simple straight-line relationship between the maximum differential entropy and the logarithm of the $L_p$-norm. Corresponding expressions for discrete and constrained continuous random variables are given parametrically; closed-form expressions are available only for special cases. However, simpler alternative bounds on the maximum entropy of integer-valued random variables are obtained by applying the differential entropy results to continuous random variables which approximate the integer-valued random variables in a natural manner.

The results tabulated here have at least two potentially useful applications. First, they can lend a theoretical underpinning to source coding distortion measures based on $L_p$-norms. Second, they can be used to perform estimates of the local entropy of a dataset, for which the available local data are sufficient for obtaining good estimates of the dataset's $L_p$-norm but not for a good estimate of its histogram. Follow-up articles on these two applications will appear in future issues.


# Acknowledgments

Much of the impetus for the work in this article came from a derivation of the $p = 2$ case of Eq. (55) in an unnumbered JPL memorandum by Bob McEliece. Mehrdad Shahshahani has pointed out that most of the results in this article remain valid for generalized $L_p$-norms with real-valued $p$, as long as $p \geq 1$. Bob McEliece has noted that similar maximum entropy expressions can be obtained when the $L_p$-norm constraint is replaced by a constraint on the expected value of an arbitrary function of the random variable.

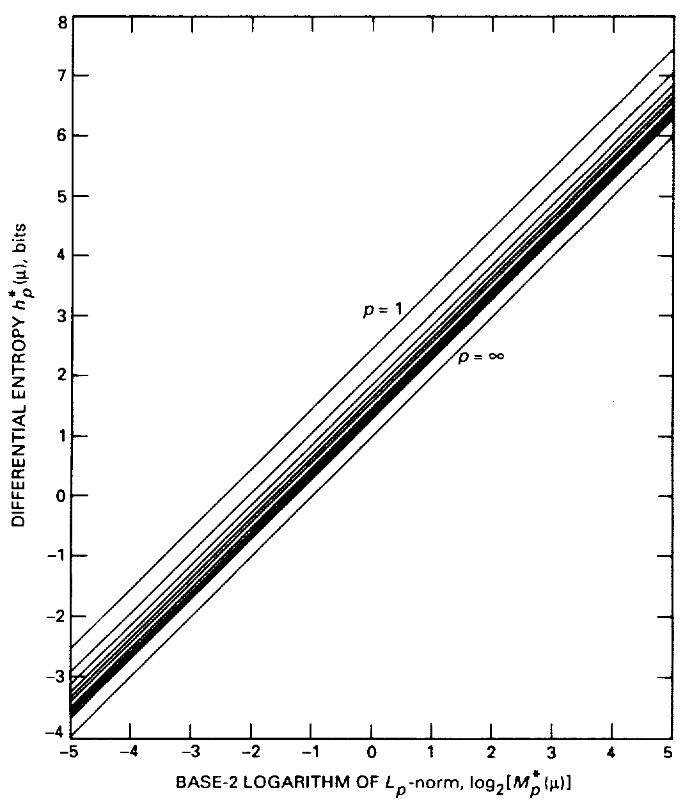


Fig. 1. Maximum differential entropy as a function of $L_p$-norm ($p = 1, 2, 3, 4, 5, 6, 8, 10, 12, 16, \infty$) for unconstrained continuous random variables.

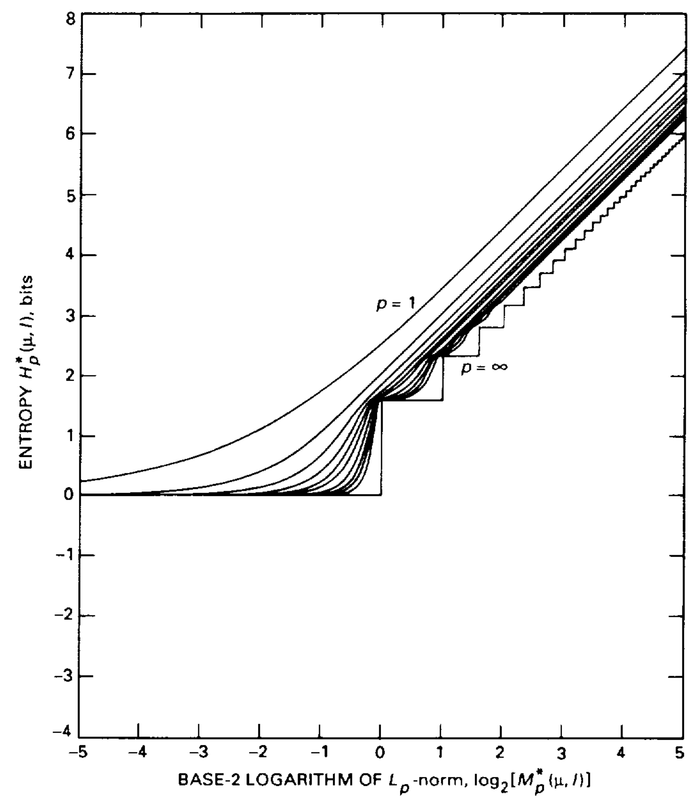


Fig. 2. Maximum entropy as a function of $L_p$-norm ($p = 1, 2, 3, 4, 5, 6, 8, 10, 12, 16, \infty$) for integer-valued random variables.

# Appendix

This appendix contains proofs or derivations omitted in the main text. Equations (1), (2), (3), (4), (9), (10), (22), (23), (25), (28), (32), (35), (36), (38), (41), (45), (48), and (49) are definitions and require no proof. Equations (7), (16), (17), (19), (20), (21), (24), (29), (30), (31), (37), (42), (43), (44), (53), (54), (55), and (56) are trivial or straightforward applications of preceding results. This leaves Eqs. (5), (6), (8), (11), (12), (13), (14), (15), (18), (26), (27), (33), (34), (39), (40), (46), (47), (50), (51), and (52) requiring further justification.

Equation (5) follows from the linearity of the expectation operator. Equations (11) and (12) come from standard integral tables [2]. Equations (13) and (15) require two elementary properties [2] of the gamma function: $\Gamma(1 + 1/p) = \Gamma(1/p)/p$ and $\Gamma(1/2) = \sqrt{\pi}$. Equations (6) and (8) result from applying the definitions in Eqs. (1) and (3) to the probability distributions of scaled and shifted random variables, obtained from standard texts [3] as:

$$\begin{aligned} f'(x') &= f(x'/q)/|q| & \qquad f''(x'') &= f(x'' + \Delta) \\ F'(X') &= F(X'/q) & \qquad F''(X'') &= F(X'' + \Delta) \end{aligned} \tag{A-1}$$

where $f'(x')$, $F'(X')$, $f''(x'')$, and $F''(X'')$ are probability density or probability mass functions for the scaled and shifted random variables $x'$, $X'$, $x''$, and $X''$.

Equations (14), (26), and (39) follow after observing that the logarithms of the probability distributions in Eqs. (10), (22), and (35) all consist of two terms, one term a constant and the second term proportional to $|x|^p$ or $|X|^p$. The expected value of the second term can thus be calculated directly from the preceding formulas, Eqs. (13), (24), and (37), for the $L_p$-norm.

Equations (18), (27), and (40) are the central results of this article and are proved by generalizing a technique used in [4] to show that maximum differential entropy with constrained second moment is achieved by a Gaussian distribution. If $x$ and $x_p^*(\mu, \Xi)$ both have $L_p$-norm $\mathbf{M}_p\{x\}$, then for $p < \infty$,

$$\begin{aligned} \mathbf{h}\{x_p^*(\mu,\Xi)\} &= -\int_\Xi f_p^*(x;\mu,\Xi)\log[f_p^*(x;\mu,\Xi)]\,dx \\ &= \int_\Xi f_p^*(x;\mu,\Xi)\left\{\log[\alpha_p^*(\mu,\Xi)] + \frac{\log[e]}{p}\,\frac{|x|^p}{\mu^p}\right\}\,dx \\ &= \int_\Xi f(x)\left\{\log[\alpha_p^*(\mu,\Xi)] + \frac{\log[e]}{p}\,\frac{|x|^p}{\mu^p}\right\}\,dx \\ &= -\int_\Xi f(x)\log[f_p^*(x;\mu,\Xi)]\,dx \end{aligned} \tag{A-2}$$

The third equality in Eq. (A-2) follows from the assumption that $x$ and $x_p^*(\mu, \Xi)$ have identical $L_p$-norms, hence $|x|^p$ has the same expectation whether it is averaged over $f_p^*(x; \mu, \Xi)$ or $f(x)$. If $p = \infty$, the same result holds: the second term in the second and third lines of Eq. (A-2) is absent, and the integration over $\Xi$ is replaced by an integration over $\Xi \cap \{|x| \le \mu\}$. Continuing,

$$\begin{aligned} \mathbf{h}\{x_p^*(\mu,\Xi)\} - \mathbf{h}\{x\} &= -\int_\Xi f(x)\log[f_p^*(x;\mu,\Xi)]\,dx + \int_{\Xi\cap\{f(x)>0\}} f(x)\log[f(x)]\,dx \\ &= \int_{\Xi\cap\{f(x)>0\}} f(x)\log\left[\frac{f(x)}{f_p^*(x;\mu,\Xi)}\right]dx \\ &\ge \int_{\Xi\cap\{f(x)>0\}} f(x)\log[e]\left\{1 - \frac{f_p^*(x;\mu,\Xi)}{f(x)}\right\}dx = 0 \end{aligned} \tag{A-3}$$

The inequality in Eq. (A-3) results from the general inequality $\log[a] \geq \log[e](1 - 1/a)$ for all $a > 0$, and the last equality arises becauses $f_p^*(x;\mu,\Xi)$ and $f(x)$ both integrate to one.

The derivation in Eqs. (A-2) and (A-3) proves Eq. (27). Equation (18) is a special case of Eq. (27) obtained by setting $\Xi$ equal to the set of all real numbers. Equation (40) is derived in a similar manner by replacing the integrals in Eqs. (A-2) and (A-3) with summations and continuous probability density functions with discrete probability mass functions.

Equation (33) results from noting that $|x_p^*(\mu, R^+)| \Leftrightarrow |x_p^*(\mu)|$, so the $L_p$-norms of $x_p^*(\mu, R^+)$ and $x_p^*(\mu)$ must be identical. Equation (34) comes from the fact that the constant scale factor $\alpha_p^*(\mu, R^+)$ for $f_p^*(x;\mu,R^+)$ in Eq. (22) with $\Xi = R^+$ is exactly half the corresponding scale factor for $f_p^*(x;\mu)$ in Eq. (10). This accounts for a difference of $\log[2]$ in the first terms in their respective expressions for differential entropy. The second terms must be equal by the previous observation linking them to their respective $L_p$-norms.

To derive Eqs. (46) and (47), let $a = e^{-1/\mu}$ and replace $\Xi$ with $I$ or $I^+$ in Eqs. (36), (37), and (38) to obtain

$$A_1^*(\mu, I) = \sum_{i=-\infty}^{\infty} e^{-|i|/\mu} = 2\sum_{i=0}^{\infty} a^i - 1 = \frac{2}{1-a} - 1 = \frac{1+a}{1-a}$$

$$\mu B_1^*(\mu, I) = \sum_{i=-\infty}^{\infty} |i| e^{-|i|/\mu} = 2\sum_{i=0}^{\infty} i a^i = \frac{2a}{(1-a)^2}$$

$$A_\infty^*(\mu, I) = \sum_{|i|\leq\mu} 1 = 2\lfloor\mu\rfloor + 1$$

$$M_\infty^*(\mu, I) = \sup_{|i|\leq\mu} |i| = \lfloor\mu\rfloor \tag{A-4}$$

and

$$A_1^*(\mu, I^+) = \sum_{i=1}^{\infty} e^{-|i|/\mu} = \sum_{i=1}^{\infty} a^i = \frac{1}{1-a} - 1 = \frac{a}{1-a}$$

$$\mu B_1^*(\mu, I^+) = \sum_{i=1}^{\infty} |i| e^{-|i|/\mu} = \sum_{i=1}^{\infty} i a^i = \frac{a}{(1-a)^2}$$

$$A_\infty^*(\mu, I^+) = \sum_{1\leq i\leq\mu} 1 = \lfloor\mu\rfloor$$

$$M_\infty^*(\mu, I^+) = \sup_{1\leq i\leq\mu} |i| = \lfloor\mu\rfloor \tag{A-5}$$

where $\lfloor\mu\rfloor$ is the integer part of $\mu$. The entropy expressions in Eqs. (46) and (47) follow algebraically upon substitution of Eqs. (A-4) and (A-5) into Eqs. (37) and (39) and solving for the entropy in terms of the corresponding $L_p$-norm.

Equation (50) results from calculating the conditional probability density $f(x|X)$ of $x$ given $X$, then averaging over $X$:

$$f(x|X) = \begin{cases} 1, & \text{if } |x - X| \leq 1/2 \\ 0, & \text{if } |x - X| > 1/2 \end{cases} = \begin{cases} 1, & \text{if } X = \lfloor x + 1/2 \rfloor \\ 0, & \text{if } X \neq \lfloor x + 1/2 \rfloor \end{cases}$$

$$f(x) = \sum_{i=-\infty}^{\infty} F(i) f(x|X = i) = F(\lfloor x + 1/2 \rfloor) \tag{A-6}$$

Equation (51) results from breaking up the defining integral in Eq. (1) for the differential entropy into a sum of integrals over unit intervals,

$$\mathbf{h}\{x\} = -\int_{-\infty}^{\infty} f(x) \log[f(x)]\, dx = -\sum_{i=-\infty}^{\infty} \int_{i-1/2}^{i+1/2} f(x) \log[f(x)]\, dx$$

$$= -\sum_{i=-\infty}^{\infty} \int_{i-1/2}^{i+1/2} F(i) \log[F(i)]\, dx = -\sum_{i=-\infty}^{\infty} F(i) \log[F(i)] = \mathbf{H}\{X\} \tag{A-7}$$

Equation (52) is derived by considering the cases of even and odd values of $p$ separately. In the first case, when $p$ is even,

$$\mathbf{E}\{|X + u|^p\} = \mathbf{E}\{(X + u)^p\} = \sum_{r=0}^{p} \binom{p}{r} \mathbf{E}\{X^{p-r}\} \mathbf{E}\{u^r\}$$

$$= \sum_{\substack{r=0 \\ r \text{ even}}}^{p} \binom{p}{r} \mathbf{E}\{X^{p-r}\} \frac{2^{-r}}{r+1} \tag{A-8}$$

because

$$\mathbf{E}\{u^r\} = \begin{cases} \dfrac{2^{-r}}{r+1}, & \text{if } r \text{ is even} \\ 0, & \text{if } r \text{ is odd} \end{cases} \tag{A-9}$$

Thus, since $X^{p-r} = |X|^{p-r}$ when $p$ and $r$ are both even,

$$\mathbf{E}\{|X + u|^p\} = \frac{2^{-p}}{p+1} + \sum_{\substack{r=0 \\ r \text{ even}}}^{p-2} \binom{p}{r} \frac{2^{-r}}{r+1} \mathbf{E}\{|X|^{p-r}\} \tag{A-10}$$

In the second case, when $p$ is odd, the derivation begins by writing

$$|X + u| = |X| + w \tag{A-11}$$

where

$$w = \begin{cases} +u, & X \geq 1 \\ |u|, & X = 0 \\ -u, & X \leq 1 \end{cases} \tag{A-12}$$

This decomposition is valid because $X$ is integer valued. The conditional moments of $w$ are

$$\mathbf{E}\{w^r | X\} = \begin{cases} \dfrac{2^{-r}}{r+1}, & \text{if } r \text{ is even or if } r \text{ is odd and } X = 0 \\ 0, & \text{if } r \text{ is odd and } X \neq 0 \end{cases} \tag{A-13}$$

Thus,

$$\mathbf{E}\{|X+u|^p\} = \mathbf{E}\{(|X| \ + \ w)^p\} = \mathbf{E}\left\{\sum_{r=0}^{p} \binom{p}{r} |X|^{p-r} w^r\right\}$$

$$= \sum_{i \neq 0} F(i) \sum_{r=0}^{p} \binom{p}{r} |i|^{p-r} \mathbf{E}\{w^r | X = i\} + F(0)\mathbf{E}\{w^p | X = 0\}$$

$$= \sum_{i \neq 0} F(i) \sum_{\substack{r=0 \\ r \text{ even}}}^{p} \binom{p}{r} |i|^{p-r} \frac{2^{-r}}{r+1} + F(0)\frac{2^{-p}}{p+1}$$

$$= \sum_{\substack{r=0 \\ r \text{ even}}}^{p-1} \binom{p}{r} \frac{2^{-r}}{r+1} \sum_{i \neq 0} F(i)|i|^{p-r} + \ F(0)\frac{2^{-p}}{p+1}$$

$$= \sum_{\substack{r=0 \\ r \text{ even}}}^{p-1} \binom{p}{r} \frac{2^{-r}}{r+1} \mathbf{E}\{|X|^{p-r}\} + \ F(0)\frac{2^{-p}}{p+1} \tag{A-14}$$

# References

[1] C. E. Shannon "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 (Part I), pp. 623–656 (Part II), 1948; reprinted in book form with postscript by W. Weaver, Urbana: University of Illinois Press, 1949.

[2] *CRC Standard Mathematical Tables*, ed. W. H. Beyer, Boca Raton, Florida: CRC Press, Inc., 1987.

[3] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill, 1965.

[4] R. E. Blahut, *Principles and Practice of Information Theory*, Reading, Massachusetts: Addison-Wesley, 1988.