alternative proof of the entropy power inequality. We also show how the entropy power inequality and the Brunn–Minkowski inequality are related by means of a common proof.

We can rewrite the entropy power inequality for dimension $n = 1$ in a form that emphasizes its relationship to the normal distribution. Let X and Y be two independent random variables with densities, and let X' and Y' be independent normals with the same entropy as X and Y, respectively. Then $2^{2h(X)} = 2^{2h(X')} = (2\pi e)\sigma^2_{X'}$ and similarly, $2^{2h(Y)} = (2\pi e)\sigma^2_{Y'}$. Hence the entropy power inequality can be rewritten as

$$2^{2h(X+Y)} \geq (2\pi e)(\sigma^2_{X'} + \sigma^2_{Y'}) = 2^{2h(X'+Y')}, \tag{17.89}$$

since X' and Y' are independent. Thus, we have a new statement of the entropy power inequality.

Theorem 17.8.1 (*Restatement of the entropy power inequality*) *For two independent random variables X and Y,*

$$h(X + Y) \geq h(X' + Y'), \tag{17.90}$$

where X' and Y' are independent normal random variables with $h(X') = h(X)$ and $h(Y') = h(Y)$.

This form of the entropy power inequality bears a striking resemblance to the Brunn–Minkowski inequality, which bounds the volume of set sums.

Definition The *set sum* $A + B$ of two sets $A, B \subset \mathcal{R}^n$ is defined as the set $\{x + y : x \in A, y \in B\}$.

Example 17.8.1 The set sum of two spheres of radius 1 is a sphere of radius 2.

Theorem 17.8.2 *(Brunn–Minkowski inequality) The volume of the set sum of two sets A and B is greater than the volume of the set sum of two spheres A' and B' with the same volume as A and B, respectively:*

$$V(A + B) \geq V(A' + B'), \tag{17.91}$$

where A' and B' are spheres with $V(A') = V(A)$ and $V(B') = V(B)$.

The similarity between the two theorems was pointed out in [104]. A common proof was found by Dembo [162] and Lieb, starting from a strengthened version of Young's inequality. The same proof can be used to prove a range of inequalities which includes the entropy power inequality and the Brunn–Minkowski inequality as special cases. We begin with a few definitions.

Definition Let f and g be two densities over $\mathcal{R}^n$ and let $f * g$ denote the convolution of the two densities. Let the $\mathcal{L}_r$ norm of the density be defined by

$$||f||_r = \left(\int f^r(x)\,dx\right)^{\frac{1}{r}}. \tag{17.92}$$

Lemma 17.8.1 (*Strengthened Young's inequality*) *For any two densities f and g over $\mathcal{R}^n$,*

$$||f * g||_r \le \left(\frac{C_p C_q}{C_r}\right)^{\frac{n}{2}} ||f||_p ||g||_q, \tag{17.93}$$

where

$$\frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1 \tag{17.94}$$

and

$$C_p = \frac{p^{\frac{1}{p}}}{p'^{\frac{1}{p'}}}, \qquad \frac{1}{p} + \frac{1}{p'} = 1. \tag{17.95}$$

Proof: The proof of this inequality may be found in [38] and [73]. □

We define a generalization of the entropy.

Definition The *Renyi entropy* $h_r(X)$ of order r is defined as

$$h_r(X) = \frac{1}{1-r}\log\left[\int f^r(x)\,dx\right] \tag{17.96}$$

for $0 < r < \infty$, $r \neq 1$. If we take the limit as $r \to 1$, we obtain the Shannon entropy function,

$$h(X) = h_1(X) = -\int f(x)\log f(x)\,dx. \tag{17.97}$$

If we take the limit as $r \to 0$, we obtain the logarithm of the volume of the support set,

$$h_0(X) = \log\left(\mu\{x : f(x) > 0\}\right). \tag{17.98}$$

Thus, the zeroth-order Renyi entropy gives the logarithm of the measure of the support set of the density f, and the Shannon entropy h_1 gives the logarithm of the size of the "effective" support set (Theorem 8.2.2). We now define the equivalent of the entropy power for Renyi entropies.

Definition The *Renyi entropy power* $V_r(X)$ of order r is defined as

$$V_r(X) = \begin{cases} \left[\int f^r(x)\,dx\right]^{-\frac{2}{n}\frac{r'}{r}}, & 0 < r \le \infty\,, r \ne 1\,, \frac{1}{r} + \frac{1}{r'} = 1 \\ \exp\left[\frac{2}{n} h(X)\right], & r = 1 \\ \mu(\{x : f(x) > 0\})^{\frac{2}{n}}, & r = 0 \end{cases} \tag{17.99}$$

Theorem 17.8.3 *For two independent random variables X and Y and any* $0 \le r < \infty$ *and any* $0 \le \lambda \le 1$, *we have*

$$\log V_r(X+Y) \ge \lambda \log V_p(X) + (1-\lambda)\log V_q(Y) + H(\lambda) + \frac{1+r}{1-r}\left[H\left(\frac{r+\lambda(1-r)}{1+r}\right) - H\left(\frac{r}{1+r}\right)\right], \tag{17.100}$$

where $p = \frac{r}{(r+\lambda(1-r))}$, $q = \frac{r}{(r+(1-\lambda)(1-r))}$ *and* $H(\lambda) = -\lambda\log\lambda - (1-\lambda)\log(1-\lambda)$.

Proof: If we take the logarithm of Young's inequality (17.93), we obtain

$$\frac{1}{r'}\log V_r(X+Y) \ge \frac{1}{p'}\log V_p(X) + \frac{1}{q'}\log V_q(Y) + \log C_r - \log C_p - \log C_q. \tag{17.101}$$

Setting $\lambda = r'/p'$ and using (17.94), we have $1-\lambda = r'/q'$, $p = \frac{r}{r+\lambda(1-r)}$ and $q = \frac{r}{r+(1-\lambda)(1-r)}$. Thus, (17.101) becomes

$$\log V_r(X+Y) \ge \lambda\log V_p(X) + (1-\lambda)\log V_q(Y) + \frac{r'}{r}\log r - \log r' - \frac{r'}{p}\log p + \frac{r'}{p'}\log p' - \frac{r'}{q}\log q + \frac{r'}{q'}\log q' \tag{17.102}$$

$$= \lambda\log V_p(X) + (1-\lambda)\log V_q(Y) + \frac{r'}{r}\log r - (\lambda + 1 - \lambda)\log r' - \frac{r'}{p}\log p + \lambda\log p' - \frac{r'}{q}\log q + (1-\lambda)\log q' \tag{17.103}$$

$$= \lambda \log V_p(X) + (1-\lambda)\log V_q(Y) + \frac{1}{r-1}\log r + H(\lambda)$$
$$- \frac{r+\lambda(1-r)}{r-1}\log\frac{r}{r+\lambda(1-r)}$$
$$- \frac{r+(1-\lambda)(1-r)}{r-1}\log\frac{r}{r+(1-\lambda)(1-r)} \tag{17.104}$$

$$= \lambda \log V_p(X) + (1-\lambda)\log V_q(Y) + H(\lambda)$$
$$+ \frac{1+r}{1-r}\left[H\left(\frac{r+\lambda(1-r)}{1+r}\right) - H\left(\frac{r}{1+r}\right)\right], \tag{17.105}$$

where the details of the algebra for the last step are omitted. □

The Brunn–Minkowski inequality and the entropy power inequality can then be obtained as special cases of this theorem.

- *The entropy power inequality.* Taking the limit of (17.100) as $r \to 1$ and setting

$$\lambda = \frac{V_1(X)}{V_1(X)+V_1(Y)}, \tag{17.106}$$

  we obtain

$$V_1(X+Y) \geq V_1(X) + V_1(Y), \tag{17.107}$$

  which is the entropy power inequality.

- *The Brunn–Minkowski inequality.* Similarly, letting $r \to 0$ and choosing

$$\lambda = \frac{\sqrt{V_0(X)}}{\sqrt{V_0(X)}+\sqrt{V_0(Y)}}, \tag{17.108}$$

  we obtain

$$\sqrt{V_0(X+Y)} \geq \sqrt{V_0(X)} + \sqrt{V_0(Y)}. \tag{17.109}$$

  Now let A be the support set of X and B be the support set of Y. Then $A+B$ is the support set of $X+Y$, and (17.109) reduces to

$$[\mu(A+B)]^{\frac{1}{n}} \geq [\mu(A)]^{\frac{1}{n}} + [\mu(B)]^{\frac{1}{n}}, \tag{17.110}$$

  which is the Brunn–Minkowski inequality.

The general theorem unifies the entropy power inequality and the Brunn–Minkowski inequality and introduces a continuum of new inequalities that lie between the entropy power inequality and the Brunn–Minkowski inequality. This further strengthens the analogy between entropy power and volume.

17.9 INEQUALITIES FOR DETERMINANTS

Throughout the remainder of this chapter, we assume that K is a nonnegative definite symmetric $n \times n$ matrix. Let $|K|$ denote the determinant of K.

We first give an information-theoretic proof of a result due to Ky Fan [199].

Theorem 17.9.1 $\log |K|$ *is concave.*

Proof: Let X_1 and X_2 be normally distributed n-vectors, $\mathbf{X}_i \sim \mathcal{N}(0, K_i)$, $i = 1, 2$. Let the random variable θ have the distribution

$$\Pr\{\theta = 1\} = \lambda, \tag{17.111}$$

$$\Pr\{\theta = 2\} = 1 - \lambda \tag{17.112}$$

for some $0 \le \lambda \le 1$. Let θ, $\mathbf{X}_1$, and $\mathbf{X}_2$ be independent, and let $\mathbf{Z} = \mathbf{X}_\theta$. Then $\mathbf{Z}$ has covariance $K_Z = \lambda K_1 + (1-\lambda)K_2$. However, $\mathbf{Z}$ will not be multivariate normal. By first using Theorem 17.2.3, followed by Theorem 17.2.1, we have

$$\frac{1}{2}\log(2\pi e)^n|\lambda K_1 + (1-\lambda)K_2| \ge h(\mathbf{Z}) \tag{17.113}$$

$$\ge h(\mathbf{Z}|\theta) \tag{17.114}$$

$$= \lambda \frac{1}{2}\log(2\pi e)^n|K_1| + (1-\lambda)\frac{1}{2}\log(2\pi e)^n|K_2|.$$

Thus,

$$|\lambda K_1 + (1-\lambda)K_2| \ge |K_1|^\lambda |K_2|^{1-\lambda}, \tag{17.115}$$

as desired. □

We now give Hadamard's inequality using an information-theoretic proof [128].

Theorem 17.9.2 (*Hadamard*) $|K| \leq \Pi K_{ii}$, *with equality iff* $K_{ij} = 0,\ i \neq j$.

Proof: Let $\mathbf{X} \sim \mathcal{N}(0, K)$. Then

$$\frac{1}{2}\log(2\pi e)^n |K| = h(X_1, X_2, \ldots, X_n) \leq \sum h(X_i) = \sum_{i=1}^{n} \frac{1}{2}\log 2\pi e|K_{ii}|, \tag{17.116}$$

with equality iff $X_1, X_2, \ldots, X_n$ are independent (i.e., $K_{ij} = 0, i \neq j$).□

We now prove a generalization of Hadamard's inequality due to Szasz [391]. Let $K(i_1, i_2, \ldots, i_k)$ be the $k \times k$ principal submatrix of K formed by the rows and columns with indices $i_1, i_2, \ldots, i_k$.

Theorem 17.9.3 (*Szasz*) *If K is a positive definite $n \times n$ matrix and P_k denotes the product of the determinants of all the principal k-rowed minors of K, that is,*

$$P_k = \prod_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} |K(i_1, i_2, \ldots, i_k)|, \tag{17.117}$$

then

$$P_1 \geq P_2^{\frac{1}{\binom{n-1}{1}}} \geq P_3^{\frac{1}{\binom{n-1}{2}}} \geq \cdots \geq P_n. \tag{17.118}$$

Proof: Let $\mathbf{X} \sim \mathcal{N}(0, K)$. Then the theorem follows directly from Theorem 17.6.1, with the identification $h_k^{(n)} = \frac{1}{2n\binom{n-1}{k-1}} \log P_k + \frac{1}{2}\log 2\pi e$. □

We can also prove a related theorem.

Theorem 17.9.4 *Let K be a positive definite $n \times n$ matrix and let*

$$S_k^{(n)} = \frac{1}{\binom{n}{k}} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} |K(i_1, i_2, \ldots, i_k)|^{\frac{1}{k}}. \tag{17.119}$$

Then

$$\frac{1}{n} tr(K) = S_1^{(n)} \geq S_2^{(n)} \geq \cdots \geq S_n^{(n)} = |K|^{\frac{1}{n}}. \tag{17.120}$$

Proof: This follows directly from the corollary to Theorem 17.6.1, with the identification $t_k^{(n)} = (2\pi e) S_k^{(n)}$ and $r = 2$. □

Theorem 17.9.5 *Let*

$$Q_k = \left(\prod_{S:|S|=k} \frac{|K|}{|K(S^c)|} \right)^{\frac{1}{k\binom{n}{k}}}. \tag{17.121}$$

Then

$$\left(\prod_{i=1}^{n} \sigma_i^2 \right)^{\frac{1}{n}} = Q_1 \le Q_2 \le \cdots \le Q_{n-1} \le Q_n = |K|^{\frac{1}{n}}. \tag{17.122}$$

Proof: The theorem follows immediately from Theorem 17.6.3 and the identification

$$h(X(S)|X(S^c)) = \frac{1}{2} \log(2\pi e)^k \frac{|K|}{|K(S^c)|}. \quad \square \tag{17.123}$$

The outermost inequality, $Q_1 \le Q_n$, can be rewritten as

$$|K| \ge \prod_{i=1}^{n} \sigma_i^2, \tag{17.124}$$

where

$$\sigma_i^2 = \frac{|K|}{|K(1, 2 \ldots, i-1, i+1, \ldots, n)|} \tag{17.125}$$

is the minimum mean-squared error in the linear prediction of X_i from the remaining X's. Thus, σ_i^2 is the conditional variance of X_i given the remaining X_j's if $X_1, X_2, \ldots, X_n$ are jointly normal. Combining this with Hadamard's inequality gives upper and lower bounds on the determinant of a positive definite matrix.

Corollary

$$\prod_i K_{ii} \ge |K| \ge \prod_i \sigma_i^2. \tag{17.126}$$

Hence, the determinant of a covariance matrix lies between the product of the unconditional variances K_{ii} of the random variables X_i and the product of the conditional variances σ_i^2.

We now prove a property of Toeplitz matrices, which are important as the covariance matrices of stationary random processes. A Toeplitz matrix K is characterized by the property that $K_{ij} = K_{rs}$ if $|i - j| = |r - s|$. Let K_k denote the principal minor $K(1, 2, \ldots, k)$. For such a matrix, the following property can be proved easily from the properties of the entropy function.

Theorem 17.9.6 *If the positive definite $n \times n$ matrix K is Toeplitz, then*

$$|K_1| \geq |K_2|^{\frac{1}{2}} \geq \cdots \geq |K_{n-1}|^{\frac{1}{(n-1)}} \geq |K_n|^{\frac{1}{n}} \tag{17.127}$$

and $|K_k|/|K_{k-1}|$ is decreasing in k, and

$$\lim_{n\to\infty} |K_n|^{\frac{1}{n}} = \lim_{n\to\infty} \frac{|K_n|}{|K_{n-1}|}. \tag{17.128}$$

Proof: Let $(X_1, X_2, \ldots, X_n) \sim \mathcal{N}(0, K_n)$. We observe that

$$h(X_k|X_{k-1}, \ldots, X_1) = h(X^k) - h(X^{k-1}) \tag{17.129}$$

$$= \frac{1}{2} \log(2\pi e) \frac{|K_k|}{|K_{k-1}|}. \tag{17.130}$$

Thus, the monotonicity of $|K_k|/|K_{k-1}|$ follows from the monotonocity of $h(X_k|X_{k-1}, \ldots, X_1)$, which follows from

$$h(X_k|X_{k-1}, \ldots, X_1) = h(X_{k+1}|X_k, \ldots, X_2) \tag{17.131}$$

$$\geq h(X_{k+1}|X_k, \ldots, X_2, X_1), \tag{17.132}$$

where the equality follows from the Toeplitz assumption and the inequality from the fact that conditioning reduces entropy. Since $h(X_k|X_{k-1}, \ldots, X_1)$ is decreasing, it follows that the running averages

$$\frac{1}{k} h(X_1, \ldots, X_k) = \frac{1}{k} \sum_{i=1}^{k} h(X_i|X_{i-1}, \ldots, X_1) \tag{17.133}$$

are decreasing in k. Then (17.127) follows from $h(X_1, X_2, \ldots, X_k) = \frac{1}{2} \log(2\pi e)^k |K_k|$. □

Finally, since $h(X_n|X_{n-1}, \ldots, X_1)$ is a decreasing sequence, it has a limit. Hence by the theorem of the Cesáro mean,

$$\lim_{n\to\infty} \frac{h(X_1, X_2, \ldots, X_n)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} h(X_k|X_{k-1}, \ldots, X_1)$$

$$= \lim_{n\to\infty} h(X_n|X_{n-1}, \ldots, X_1). \tag{17.134}$$

Translating this to determinants, one obtains

$$\lim_{n\to\infty} |K_n|^{\frac{1}{n}} = \lim_{n\to\infty} \frac{|K_n|}{|K_{n-1}|}. \tag{17.135}$$

Theorem 17.9.7 (*Minkowski inequality [390]*)

$$|K_1 + K_2|^{1/n} \ge |K_1|^{1/n} + |K_2|^{1/n}. \tag{17.136}$$

Proof: Let $\mathbf{X}_1, \mathbf{X}_2$ be independent with $\mathbf{X}_i \sim \mathcal{N}(0, K_i)$. Noting that $\mathbf{X}_1 + \mathbf{X}_2 \sim \mathcal{N}(0, K_1 + K_2)$ and using the entropy power inequality (Theorem 17.7.3) yields

$$(2\pi e)|K_1 + K_2|^{1/n} = 2^{\frac{2}{n}h(\mathbf{X}_1 + \mathbf{X}_2)} \tag{17.137}$$

$$\ge 2^{\frac{2}{n}h(\mathbf{X}_1)} + 2^{\frac{2}{n}h(\mathbf{X}_2)} \tag{17.138}$$

$$= (2\pi e)|K_1|^{1/n} + (2\pi e)|K_2|^{1/n}. \quad \square \tag{17.139}$$

17.10 INEQUALITIES FOR RATIOS OF DETERMINANTS

We now prove similar inequalities for ratios of determinants. Before developing the next theorem, we make an observation about minimum mean-squared-error linear prediction. If $(X_1, X_2, \ldots, X_n) \sim \mathcal{N}(0, K_n)$, we know that the conditional density of X_n given $(X_1, X_2, \ldots, X_{n-1})$ is univariate normal with mean linear in $X_1, X_2, \ldots, X_{n-1}$ and conditional variance σ_n^2. Here σ_n^2 is the minimum mean squared error $E(X_n - \hat{X}_n)^2$ over all linear estimators $\hat{X}_n$ based on $X_1, X_2, \ldots, X_{n-1}$.

Lemma 17.10.1 $\sigma_n^2 = |K_n|/|K_{n-1}|$.

Proof: Using the conditional normality of X_n, we have

$$\frac{1}{2}\log 2\pi e\sigma_n^2 = h(X_n|X_1, X_2, \ldots, X_{n-1}) \tag{17.140}$$

$$= h(X_1, X_2, \ldots, X_n) - h(X_1, X_2, \ldots, X_{n-1}) \tag{17.141}$$

$$= \frac{1}{2}\log(2\pi e)^n|K_n| - \frac{1}{2}\log(2\pi e)^{n-1}|K_{n-1}| \tag{17.142}$$

$$= \frac{1}{2}\log 2\pi e|K_n|/|K_{n-1}|. \quad \square \tag{17.143}$$

Minimization of σ_n^2 over a set of allowed covariance matrices $\{K_n\}$ is aided by the following theorem. Such problems arise in maximum entropy spectral density estimation.

Theorem 17.10.1 *(Bergstrøm [42])* $\log(|K_n|/|K_{n-p}|)$ *is concave in* K_n.

Proof: We remark that Theorem 17.9.1 cannot be used because $\log(|K_n|/|K_{n-p}|)$ is the difference of two concave functions. Let $\mathbf{Z} = \mathbf{X}_\theta$, where $\mathbf{X}_1 \sim \mathcal{N}(0, S_n)$, $\mathbf{X}_2 \sim \mathcal{N}(0, T_n)$, $\Pr\{\theta = 1\} = \lambda = 1 - \Pr\{\theta = 2\}$, and let $\mathbf{X}_1, \mathbf{X}_2, \theta$ be independent. The covariance matrix K_n of $\mathbf{Z}$ is given by

$$K_n = \lambda S_n + (1-\lambda)T_n. \tag{17.144}$$

The following chain of inequalities proves the theorem:

$$\lambda\frac{1}{2}\log(2\pi e)^p|S_n|/|S_{n-p}| + (1-\lambda)\frac{1}{2}\log(2\pi e)^p|T_n|/|T_{n-p}|$$

$$\stackrel{\text{(a)}}{=} \lambda h(X_{1,n}, X_{1,n-1}, \ldots, X_{1,n-p+1}|X_{1,1}, \ldots, X_{1,n-p}) + (1-\lambda)h(X_{2,n}, X_{2,n-1}, \ldots, X_{2,n-p+1}|X_{2,1}, \ldots, X_{2,n-p}) \tag{17.145}$$

$$= h(Z_n, Z_{n-1}, \ldots, Z_{n-p+1}|Z_1, \ldots, Z_{n-p}, \theta) \tag{17.146}$$

$$\stackrel{\text{(b)}}{\le} h(Z_n, Z_{n-1}, \ldots, Z_{n-p+1}|Z_1, \ldots, Z_{n-p}) \tag{17.147}$$

$$\stackrel{\text{(c)}}{\le} \frac{1}{2}\log(2\pi e)^p\frac{|K_n|}{|K_{n-p}|}, \tag{17.148}$$

where (a) follows from $h(X_n, X_{n-1}, \ldots, X_{n-p+1}|X_1, \ldots, X_{n-p}) = h(X_1, \ldots, X_n) - h(X_1, \ldots, X_{n-p})$, (b) follows from the conditioning lemma, and (c) follows from a conditional version of Theorem 17.2.3. □

Theorem 17.10.2 *(Bergstrøm [42])* $|K_n|/|K_{n-1}|$ *is concave in* K_n.

Proof: Again we use the properties of Gaussian random variables. Let us assume that we have two independent Gaussian random n-vectors, $\mathbf{X} \sim \mathcal{N}(0, A_n)$ and $\mathbf{Y} \sim \mathcal{N}(0, B_n)$. Let $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$. Then

$$\frac{1}{2}\log 2\pi e\frac{|A_n + B_n|}{|A_{n-1} + B_{n-1}|} \stackrel{\text{(a)}}{=} h(Z_n|Z_{n-1}, Z_{n-2}, \ldots, Z_1) \tag{17.149}$$

$$\stackrel{(b)}{\geq} h(Z_n|Z_{n-1}, Z_{n-2}, \ldots, Z_1, X_{n-1}, X_{n-2}, \ldots, X_1, Y_{n-1}, Y_{n-2}, \ldots, Y_1) \tag{17.150}$$

$$\stackrel{(c)}{=} h(X_n + Y_n|X_{n-1}, X_{n-2}, \ldots, X_1, Y_{n-1}, Y_{n-2}, \ldots, Y_1) \tag{17.151}$$

$$\stackrel{(d)}{=} E\frac{1}{2}\log\left[2\pi e \operatorname{Var}(X_n + Y_n|X_{n-1}, X_{n-2}, \ldots, X_1, Y_{n-1}, Y_{n-2}, \ldots, Y_1)\right] \tag{17.152}$$

$$\stackrel{(e)}{=} E\frac{1}{2}\log\left[2\pi e(\operatorname{Var}(X_n|X_{n-1}, X_{n-2}, \ldots, X_1) + \operatorname{Var}(Y_n|Y_{n-1}, Y_{n-2}, \ldots, Y_1))\right] \tag{17.153}$$

$$\stackrel{(f)}{=} E\frac{1}{2}\log\left(2\pi e\left(\frac{|A_n|}{|A_{n-1}|} + \frac{|B_n|}{|B_{n-1}|}\right)\right) \tag{17.154}$$

$$= \frac{1}{2}\log\left(2\pi e\left(\frac{|A_n|}{|A_{n-1}|} + \frac{|B_n|}{|B_{n-1}|}\right)\right), \tag{17.155}$$

where

(a) follows from Lemma 17.10.1
(b) follows from the fact that the conditioning decreases entropy
(c) follows from the fact that Z is a function of X and Y
(d) follows since $X_n + Y_n$ is Gaussian conditioned on $X_1, X_2, \ldots, X_{n-1}, Y_1, Y_2, \ldots, Y_{n-1}$, and hence we can express its entropy in terms of its variance
(e) follows from the independence of X_n and Y_n conditioned on the past $X_1, X_2, \ldots, X_{n-1}, Y_1, Y_2, \ldots, Y_{n-1}$

(f) follows from the fact that for a set of jointly Gaussian random variables, the conditional variance is constant, independent of the conditioning variables (Lemma 17.10.1)

Setting $A = \lambda S$ and $B = \overline{\lambda} T$, we obtain

$$\frac{|\lambda S_n + \overline{\lambda} T_n|}{|\lambda S_{n-1} + \overline{\lambda} T_{n-1}|} \geq \lambda \frac{|S_n|}{|S_{n-1}|} + \overline{\lambda} \frac{|T_n|}{|T_{n-1}|} \tag{17.156}$$

(i.e., $|K_n|/|K_{n-1}|$ is concave). Simple examples show that $|K_n|/|K_{n-p}|$ is not necessarily concave for $p \geq 2$. □

A number of other determinant inequalities can be proved by these techniques. A few of them are given as problems.

OVERALL SUMMARY

Entropy. $H(X) = -\sum p(x) \log p(x)$.

Relative entropy. $D(p||q) = \sum p(x) \log \frac{p(x)}{q(x)}$.

Mutual information. $I(X; Y) = \sum p(x, y) \log \frac{p(x,y)}{p(x)p(y)}$.

Information inequality. $D(p||q) \geq 0$.

Asymptotic equipartition property. $-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \rightarrow H(\mathcal{X})$.

Data compression. $H(X) \leq L^* < H(X) + 1$.

Kolmogorov complexity. $K(x) = \min_{\mathcal{U}(p)=x} l(p)$.

Universal probability. $\log \frac{1}{P_{\mathcal{U}}(x)} \approx K(x)$.

Channel capacity. $C = \max_{p(x)} I(X; Y)$.

Data transmission

- $R < C$: Asymptotically error-free communication possible
- $R > C$: Asymptotically error-free communication not possible

Gaussian channel capacity. $C = \frac{1}{2} \log(1 + \frac{P}{N})$.

Rate distortion. $R(D) = \min I(X; \hat{X})$ over all $p(\hat{x}|x)$ such that $E_{p(x)p(\hat{x}|x)} d(X, \hat{X}) \leq D$.

Growth rate for investment. $W^* = \max_{\mathbf{b}^*} E \log \mathbf{b}^t \mathbf{X}$.

PROBLEMS

17.1 *Sum of positive definite matrices*. For any two positive definite matrices, K_1 and K_2, show that $|K_1 + K_2| \geq |K_1|$.

17.2 *Fan's inequality [200] for ratios of determinants.* For all $1 \le p \le n$, for a positive definite $K = K(1, 2, \ldots, n)$, show that

$$\frac{|K|}{|K(p+1, p+2, \ldots, n)|} \le \prod_{i=1}^{p} \frac{|K(i, p+1, p+2, \ldots, n)|}{|K(p+1, p+2, \ldots, n)|}. \tag{17.157}$$

17.3 *Convexity of determinant ratios.* For positive definite matrices K, K_0, show that $\ln(|K + K_0|/|K|)$ is convex in K.

17.4 *Data-processing inequality.* Let random variable X_1, X_2, X_3, and X_4 form a Markov chain $X_1 \to X_2 \to X_3 \to X_4$. Show that

$$I(X_1; X_3) + I(X_2; X_4) \le I(X_1; X_4) + I(X_2; X_3). \tag{17.158}$$

17.5 *Markov chains.* Let random variables X, Y, Z, and W form a Markov chain so that $X \to Y \to (Z, W)$ [i.e., $p(x, y, z, w) = p(x)p(y|x)p(z, w|y)$]. Show that

$$I(X; Z) + I(X; W) \le I(X; Y) + I(Z; W). \tag{17.159}$$

HISTORICAL NOTES

The entropy power inequality was stated by Shannon [472]; the first formal proofs are due to Stam [505] and Blachman [61]. The unified proof of the entropy power and Brunn–Minkowski inequalities is in Dembo et al.[164].

Most of the matrix inequalities in this chapter were derived using information-theoretic methods by Cover and Thomas [118]. Some of the subset inequalities for entropy rates may be found in Han [270].

BIBLIOGRAPHY

[1] J. Abrahams. Code and parse trees for lossless source encoding. *Proc. Compression and Complexity of Sequences 1997*, pages 145–171, 1998.

[2] N. Abramson. The ALOHA system—another alternative for computer communications. *AFIPS Conf. Proc.*, pages 281–285, 1970.

[3] N. M. Abramson. *Information Theory and Coding*. McGraw-Hill, New York, 1963.

[4] Y. S. Abu-Mostafa. Information theory. *Complexity*, pages 25–28, Nov. 1989.

[5] R. L. Adler, D. Coppersmith, and M. Hassner. Algorithms for sliding block codes: an application of symbolic dynamics to information theory. *IEEE Trans. Inf. Theory*, IT-29(1):5–22, 1983.

[6] R. Ahlswede. The capacity of a channel with arbitrary varying Gaussian channel probability functions. *Trans. 6th Prague Conf. Inf. Theory*, pages 13–21, Sept. 1971.

[7] R. Ahlswede. Multi-way communication channels. In *Proc. 2nd Int. Symp. Inf. Theory* (*Tsahkadsor, Armenian S.S.R.*), pages 23–52. Hungarian Academy of Sciences, Budapest, 1971.

[8] R. Ahlswede. The capacity region of a channel with two senders and two receivers. *Ann. Prob.*, 2:805–814, 1974.

[9] R. Ahlswede. Elimination of correlation in random codes for arbitrarily varying channels. *Z. Wahrscheinlichkeitstheorie und verwandte Gebiete*, 33:159–175, 1978.

[10] R. Ahlswede. Coloring hypergraphs: A new approach to multiuser source coding. *J. Comb. Inf. Syst. Sci.*, pages 220–268, 1979.

[11] R. Ahlswede. A method of coding and an application to arbitrarily varying channels. *J. Comb. Inf. Syst. Sci.*, pages 10–35, 1980.

[12] R. Ahlswede and T. S. Han. On source coding with side information via a multiple access channel and related problems in multi-user information theory. *IEEE Trans. Inf. Theory*, IT-29:396–412, 1983.

Elements of Information Theory, Second Edition, By Thomas M. Cover and Joy A. Thomas

[13] R. Ahlswede and J. Körner. Source coding with side information and a converse for the degraded broadcast channel. *IEEE Trans. Inf. Theory*, IT-21:629–637, 1975.

[14] R. F. Ahlswede. Arbitrarily varying channels with states sequence known to the sender. *IEEE Trans. Inf. Theory*, pages 621–629, Sept. 1986.

[15] R. F. Ahlswede. The maximal error capacity of arbitrarily varying channels for constant list sizes (corresp.). *IEEE Trans. Inf. Theory*, pages 1416–1417, July 1993.

[16] R. F. Ahlswede and G. Dueck. Identification in the presence of feedback: a discovery of new capacity formulas. *IEEE Trans. Inf. Theory*, pages 30–36, Jan. 1989.

[17] R. F. Ahlswede and G. Dueck. Identification via channels. *IEEE Trans. Inf. Theory*, pages 15–29, Jan. 1989.

[18] R. F. Ahlswede, E. H. Yang, and Z. Zhang. Identification via compressed data. *IEEE Trans. Inf. Theory*, pages 48–70, Jan. 1997.

[19] H. Akaike. Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. Inf. Theory*, pages 267–281, 1973.

[20] P. Algoet and T. M. Cover. A sandwich proof of the Shannon–McMillan–Breiman theorem. *Ann. Prob.*, 16(2):899–909, 1988.

[21] P. Algoet and T. M. Cover. Asymptotic optimality and asymptotic equipartition property of log-optimal investment. *Ann. Prob.*, 16(2):876–898, 1988.

[22] S. Amari. *Differential-Geometrical Methods in Statistics*. Springer-Verlag, New York, 1985.

[23] S. I. Amari and H. Nagaoka. *Methods of Information Geometry*. Oxford University Press, Oxford, 1999.

[24] V. Anantharam and S. Verdu. Bits through queues. *IEEE Trans. Inf. Theory*, pages 4–18, Jan. 1996.

[25] S. Arimoto. An algorithm for calculating the capacity of an arbitrary discrete memoryless channel. *IEEE Trans. Inf. Theory*, IT-18:14–20, 1972.

[26] S. Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Trans. Inf. Theory*, IT-19:357–359, 1973.

[27] R. B. Ash. *Information Theory*. Interscience, New York, 1965.

[28] J. Aczél and Z. Daróczy. *On Measures of Information and Their Characterization*. Academic Press, New York, 1975.

[29] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv. Optimal decoding of linear codes for minimizing symbol error rate (corresp.). *IEEE Trans. Inf. Theory*, pages 284–287, March 1974.

[30] A. Barron. Entropy and the central limit theorem. *Ann. Prob.*, 14(1): 336–342, 1986.

[31] A. Barron and T. M. Cover. A bound on the financial value of information. *IEEE Trans. Inf. Theory*, IT-34:1097–1100, 1988.

[32] A. Barron and T. M. Cover. Minimum complexity density estimation. *IEEE Trans. Inf. Theory*, 37(4):1034–1054, July 1991.

[33] A. R. Barron. *Logically smooth density estimation*. Ph.D. thesis, Department of Electrical Engineering, Stanford University, Stanford, CA, 1985.

[34] A. R. Barron. The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman theorem. *Ann. Prob.*, 13:1292–1303, 1985.

[35] A. R. Barron. Are Bayes' rules consistent in information? *Prob. Commun. Computation*, pages 85–91, 1987.

[36] A. R. Barron, J. Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inf. Theory*, pages 2743–2760, Oct. 1998.

[37] E. B. Baum. Neural net algorithms that learn in polynomial time from examples and queries. *IEEE Trans. Neural Networks*, pages 5–19, 1991.

[38] W. Beckner. Inequalities in Fourier analysis. *Ann. Math.*, 102:159–182, 1975.

[39] R. Bell and T. M. Cover. Competitive optimality of logarithmic investment. *Math. Oper. Res.*, 5(2):161–166, May 1980.

[40] R. Bell and T. M. Cover. Game-theoretic optimal portfolios. *Manage. Sci.*, 34(6):724–733, 1988.

[41] T. C. Bell, J. G. Cleary, and I. H. Witten. *Text Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1990.

[42] R. Bellman. Notes on matrix theory. IV: An inequality due to Bergstrøm. *Am. Math. Monthly*, 62:172–173, 1955.

[43] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. *Proc. IEEE Int. Conf. Comput.*, pages 175–179, 1984.

[44] C. H. Bennett, D. P. DiVincenzo, J. Smolin, and W. K. Wootters. Mixed state entanglement and quantum error correction. *Phys. Rev. A*, pages 3824–3851, 1996.

[45] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin. Capacities of quantum erasure channels. *Phys. Rev. Lett.*, pages 3217–3220, 1997.

[46] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein–podolsky–Rosen states. *Phys. Rev. Lett.*, pages 2881–2884, 1992.

[47] C. H. Bennett. Demons, engines and the second law. *Sci. Am.*, 259(5):108–116, Nov. 1987.

[48] C. H. Bennett and R. Landauer. The fundamental physical limits of computation. *Sci. Am.*, 255(1):48–56, July 1985.

[49] C. H. Bennett and P. W. Shor. Quantum information theory. *IEEE Trans. Inf. Theory*, IT-44:2724–2742, Oct. 1998.

[50] J. Bentley, D. Sleator, R. Tarjan, and V. Wei. Locally adaptive data compression scheme. *Commun. ACM*, pages 320–330, 1986.

[51] R. Benzel. The capacity region of a class of discrete additive degraded interference channels. *IEEE Trans. Inf. Theory*, IT-25:228–231, 1979.

[52] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

[53] T. Berger. Multiterminal source coding. In G. Longo (Ed.), *The Information Theory Approach to Communications*. Springer-Verlag, New York, 1977.

[54] T. Berger and R. W. Yeung. Multiterminal source encoding with one distortion criterion. *IEEE Trans. Inf. Theory*, IT-35:228–236, 1989.

[55] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inf. Theory*, IT-19:197–207, 1973.

[56] E. R. Berlekamp. *Block Coding with Noiseless Feedback*. Ph.D. thesis, MIT, Cambridge, MA, 1964.

[57] C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo codes. *Proc. 1993 Int. Conf. Commun.*, pages 1064–1070, May 1993.

[58] D. Bertsekas and R. Gallager. *Data Networks, 2nd ed.*. Prentice-Hall, Englewood Cliffs, NJ, 1992.

[59] M. Bierbaum and H. M. Wallmeier. A note on the capacity region of the multiple access channel. *IEEE Trans. Inf. Theory*, IT-25:484, 1979.

[60] E. Biglieri, J. Proakis, and S. Shamai. Fading channels: information-theoretic and communications aspects. *IEEE Trans. Inf. Theory*, pages 2619–2692, October 1998.

[61] N. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Inf. Theory*, IT-11:267–271, Apr. 1965.

[62] D. Blackwell, L. Breiman, and A. J. Thomasian. Proof of Shannon's transmission theorem for finite-state indecomposable channels. *Ann. Math. Stat.*, pages 1209–1220, 1958.

[63] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacity of a class of channels. *Ann. Math. Stat.*, 30:1229–1241, 1959.

[64] D. Blackwell, L. Breiman, and A. J. Thomasian. The capacities of certain channel classes under random coding. *Ann. Math. Stat.*, 31:558–567, 1960.

[65] R. Blahut. Computation of channel capacity and rate distortion functions. *IEEE Trans. Inf. Theory*, IT-18:460–473, 1972.

[66] R. E. Blahut. Information bounds of the Fano–Kullback type. *IEEE Trans. Inf. Theory*, IT-22:410–421, 1976.

[67] R. E. Blahut. *Principles and Practice of Information Theory*. Addison-Wesley, Reading, MA, 1987.

[68] R. E. Blahut. Hypothesis testing and information theory. *IEEE Trans. Inf. Theory*, IT-20:405–417, 1974.

[69] R. E. Blahut. *Theory and Practice of Error Control Codes*. Addison-Wesley, Reading, MA, 1983.

[70] B. M. Hochwald, G. Caire, B. Hassibi, and T. L. Marzetta (Eds.). *IEEE Trans. Inf. Theory*, Special Issue on Space-Time Transmission, Reception, Coding and Signal-Processing, Vol. 49, Oct. 2003.

[71] L. Boltzmann. Beziehung Zwischen dem zweiten Hauptsatze der mechanischen Wärmertheorie und der Wahrscheilichkeitsrechnung respektive den Saetzen uber das Wärmegleichgwicht. *Wien. Ber.*, pages 373–435, 1877.

[72] R. C. Bose and D. K. Ray-Chaudhuri. On a class of error correcting binary group codes. *Inf. Control*, 3:68–79, Mar. 1960.

[73] H. J. Brascamp and E. J. Lieb. Best constants in Young's inequality, its converse and its generalization to more than three functions. *Adv. Math.*, 20:151–173, 1976.

[74] L. Breiman. The individual ergodic theorems of information theory. *Ann. Math. Stat.*, 28:809–811, 1957. With correction made in 31:809-810.

[75] L. Breiman. Optimal gambling systems for favourable games. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pages 65–78. University of California Press, Berkeley, CA, 1961.

[76] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees.* Wadsworth & Brooks, Pacific Grove, CA, 1984.

[77] L. Brillouin. *Science and Information Theory*. Academic Press, New York, 1962.

[78] J. A. Bucklew. The source coding theorem via Sanov's theorem. *IEEE Trans. Inf. Theory*, pages 907–909, Nov. 1987.

[79] J. A. Bucklew. *Large Deviation Techniques in Decision, Simulation, and Estimation*. Wiley, New York, 1990.

[80] J. P. Burg. *Maximum entropy spectral analysis*. Ph.D. thesis, Department of Geophysics, Stanford University, Stanford, CA, 1975.

[81] M. Burrows and D. J. Wheeler. *A Block-Sorting Lossless Data Compression Algorithm* (Tech. Rept. 124). Digital Systems Research Center, Palo Alto, CA, May 1994.

[82] A. R. Calderbank. The art of signaling: fifty years of coding theory. *IEEE Trans. Inf. Theory*, pages 2561–2595, Oct. 1998.

[83] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, pages 1098–1106, 1995.

[84] A. Carleial. Outer bounds on the capacity of the interference channel. *IEEE Trans. Inf. Theory*, IT-29:602–606, 1983.

[85] A. B. Carleial. A case where interference does not reduce capacity. *IEEE Trans. Inf. Theory*, IT-21:569–570, 1975.

[86] G. Chaitin. *Information-Theoretic Incompleteness*. World Scientific, Singapore, 1992.

[87] G. J. Chaitin. On the length of programs for computing binary sequences. *J. ACM*, pages 547–569, 1966.

[88] G. J. Chaitin. The limits of mathematics. *J. Universal Comput. Sci.*, 2(5):270–305, 1996.

[89] G. J. Chaitin. On the length of programs for computing binary sequences. *J. ACM*, 13:547–569, 1966.

[90] G. J. Chaitin. Information theoretical limitations of formal systems. *J. ACM*, 21:403–424, 1974.

[91] G. J. Chaitin. Randomness and mathematical proof. *Sci. Am.*, 232(5):47–52, May 1975.

[92] G. J. Chaitin. Algorithmic information theory. *IBM J. Res. Dev.*, 21:350–359, 1977.

[93] G. J. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, Cambridge, 1987.

[94] C. S. Chang and J. A. Thomas. Huffman algebras for independent random variables. *Discrete Event Dynam. Syst.*, 4:23–40, 1994.

[95] C. S. Chang and J. A. Thomas. Effective bandwidth in high speed digital networks. *IEEE J. Select. Areas Commun.*, 13:1091–1114, Aug. 1995.

[96] R. Chellappa. *Markov Random Fields: Theory and Applications*. Academic Press, San Diego, CA, 1993.

[97] H. Chernoff. A measure of the asymptotic efficiency of tests of a hypothesis based on a sum of observations. *Ann. Math. Stat.*, 23:493–507, 1952.

[98] B. S. Choi and T. M. Cover. An information-theoretic proof of Burg's maximum entropy spectrum. *Proc. IEEE*, 72:1094–1095, 1984.

[99] N. Chomsky. Three models for the description of language. *IEEE Trans. Inf. Theory*, pages 113–124, Sept. 1956.

[100] P. A. Chou, M. Effros, and R. M. Gray. A vector quantization approach to universal noiseless coding and quantization. *IEEE Trans. Inf. Theory*, pages 1109–1138, July 1996.

[101] K. L. Chung. A note on the ergodic theorem of information theory. *Ann. Math. Stat.*, 32:612–614, 1961.

[102] B. S. Clarke and A. R. Barron. Information-theoretic asymptotics of Bayes' methods. *IEEE Trans. Inf. Theory*, pages 453–471, May 1990.

[103] B. S. Clarke and A. R. Barron. Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Stat. Planning Inf.*, pages 37–60, Aug. 1994.

[104] M. Costa and T. M. Cover. On the similarity of the entropy power inequality and the Brunn–Minkowski inequality. *IEEE Trans. Inf. Theory*, IT-30:837–839, 1984.

[105] M. H. M. Costa. On the Gaussian interference channel. *IEEE Trans. Inf. Theory*, pages 607–615, Sept. 1985.

[106] M. H. M. Costa and A. A. El Gamal. The capacity region of the discrete memoryless interference channel with strong interference. *IEEE Trans. Inf. Theory*, pages 710–711, Sept. 1987.

[107] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications to pattern recognition. *IEEE Trans. Electron. Computation*, pages 326–334, 1965.

[108] T. M. Cover. *Universal Gambling Schemes and the Complexity Measures of Kolmogorov and Chaitin* (Tech. Rept. 12). Department of Statistics, Stanford University, Stanford, CA, Oct. 1974.

[109] T. M. Cover. Open problems in information theory. *Proc. Moscow Inf. Theory Workshop*, pages 35–36, 1975.

[110] T. M. Cover. Universal portfolios. *Math. Finance*, pages 1–29, Jan. 1991.

[111] T. M. Cover. Comments on broadcast channels. *IEEE Trans. Inf. Theory*, pages 2524–2530, Oct. 1998.

[112] T. M. Cover. *Shannon and investment*. *IEEE Inf. Theory Newslett* (Special Golden Jubilee Issue), pp. 10–11, June 1998.

[113] T. M. Cover and M. S. Chiang. Duality between channel capacity and rate distortion with two-sided state information. *IEEE Trans. Inf. Theory*, IT-48(6):1629–1638, June 2002.

[114] T. M. Cover, P. Gács, and R. M. Gray. Kolmogorov's contributions to information theory and algorithmic complexity. *Ann. Prob.*, pages 840–865, July 1989.

[115] T. M. Cover, A. A. El Gamal, and M. Salehi. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inf. Theory*, pages 648–657, Nov. 1980.

[116] T. M. Cover and P. E. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory*, pages 21–27, Jan. 1967.

[117] T. M. Cover and S. Pombra. Gaussian feedback capacity. *IEEE Trans. Inf. Theory*, pages 37–43, January 1989.

[118] T. M. Cover and J. A. Thomas. Determinant inequalities via information theory. *SIAM J. Matrix Anal. and Its Applications*, 9(3):384–392, July 1988.

[119] T. M. Cover. Broadcast channels. *IEEE Trans. Inf. Theory*, IT-18:2–14, 1972.

[120] T. M. Cover. Enumerative source encoding. *IEEE Trans. Inf. Theory*, IT-19(1):73–77, Jan. 1973.

[121] T. M. Cover. An achievable rate region for the broadcast channel. *IEEE Trans. Inf. Theory*, IT-21:399–404, 1975.

[122] T. M. Cover. A proof of the data compression theorem of Slepian and Wolf for ergodic sources. *IEEE Trans. Inf. Theory*, IT-22:226–228, 1975.

[123] T. M. Cover. An algorithm for maximizing expected log investment return. *IEEE Trans. Inf. Theory*, IT-30(2):369–373, 1984.

[124] T. M. Cover. Kolmogorov complexity, data compression and inference. In J. Skwirzynski (Ed.), *The Impact of Processing Techniques on Communications*, Vol. 91 of *Applied Sciences*. Martinus-Nijhoff, Dordrecht, The Netherlands, 1985.

[125] T. M. Cover. On the competitive optimality of Huffman codes. *IEEE Trans. Inf. Theory*, 37(1):172–174, Jan. 1991.

[126] T. M. Cover. Universal portfolios. *Math. Finance*, pages 1–29, Jan. 1991.

[127] T. M. Cover and A El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, IT-25:572–584, 1979.

[128] T. M. Cover and A. El Gamal. An information theoretic proof of Hadamard's inequality. *IEEE Trans. Inf. Theory*, IT-29(6):930–931, Nov. 1983.

[129] T. M. Cover, A. El Gamal, and M. Salehi. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inf. Theory*, IT-26:648–657, 1980.

[130] T. M. Cover. Pick the largest number, *Open Problems in Communication and Computation*. Ed. by T. M. Cover and B. Gopinath, page 152, New York, 1987.

[131] T. M. Cover and R. King. A convergent gambling estimate of the entropy of English. *IEEE Trans. Inf. Theory*, IT-24:413–421, 1978.

[132] T. M. Cover and C. S. K. Leung. Some equivalences between Shannon entropy and Kolmogorov complexity. *IEEE Trans. Inf. Theory*, IT-24:331–338, 1978.

[133] T. M. Cover and C. S. K. Leung. An achievable rate region for the multiple access channel with feedback. *IEEE Trans. Inf. Theory*, IT-27:292–298, 1981.

[134] T. M. Cover, R. J. McEliece, and E. Posner. Asynchronous multiple access channel capacity. *IEEE Trans. Inf. Theory*, IT-27:409–413, 1981.

[135] T. M. Cover and E. Ordentlich. Universal portfolios with side information. *IEEE Trans. Inf. Theory*, IT-42:348–363, Mar. 1996.

[136] T. M. Cover and S. Pombra. Gaussian feedback capacity. *IEEE Trans. Inf. Theory*, IT-35:37–43, 1989.

[137] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, NJ, 1946.

[138] I. Csiszár. Information type measures of difference of probability distributions and indirect observations. *Stud. Sci. Math. Hung.*, 2:299–318, 1967.

[139] I Csiszár. On the computation of rate distortion functions. *IEEE Trans. Inf. Theory*, IT-20:122–124, 1974.

[140] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Prob.*, pages 146–158, Feb. 1975.

[141] I Csiszár. Sanov property, generalized I-projection and a conditional limit theorem. *Ann. Prob.*, 12:768–793, 1984.

[142] I. Csiszár. Information theory and ergodic theory. *Probl. Contr. Inf. Theory*, pages 3–27, 1987.

[143] I. Csiszár. A geometric interpretation of Darroch and Ratcliff's generalized iterative scaling. *Ann. Stat.*, pages 1409–1413, 1989.

[144] I. Csiszár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.*, pages 2032–2066, Dec. 1991.

[145] I. Csiszár. Arbitrarily varying channels with general alphabets and states. *IEEE Trans. Inf. Theory*, pages 1725–1742, Nov. 1992.

[146] I. Csiszár. The method of types. *IEEE Trans. Inf. Theory*, pages 2505–2523, October 1998.

[147] I. Csiszár, T. M. Cover, and B. S. Choi. Conditional limit theorems under Markov conditioning. *IEEE Trans. Inf. Theory*, IT-33:788–801, 1987.

[148] I. Csiszár and J. Körner. Towards a general theory of source networks. *IEEE Trans. Inf. Theory*, IT-26:155–165, 1980.

[149] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.

[150] I. Csiszár and J. Körner. Feedback does not affect the reliability function of a DMC at rates above capacity (corresp.). *IEEE Trans. Inf. Theory*, pages 92–93, Jan. 1982.

[151] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory*, pages 339–348, May 1978.

[152] I. Csiszár and J. Körner. Graph decomposition: a new key to coding theorems. *IEEE Trans. Inf. Theory*, pages 5–12, Jan. 1981.

[153] I. Csiszár and G. Longo. *On the Error Exponent for Source Coding and for Testing Simple Statistical Hypotheses*. Hungarian Academy of Sciences, Budapest, 1971.

[154] I. Csiszár and P. Narayan. Capacity of the Gaussian arbitrarily varying channel. *IEEE Trans. Inf. Theory*, pages 18–26, Jan. 1991.

[155] I. Csiszár and G. Tusnády. Information geometry and alternating minimization procedures. *Statistics and Decisions*, Supplement Issue 1:205–237, 1984.

[156] G. B. Dantzig and D. R. Fulkerson. On the max-flow min-cut theorem of networks. In H. W. Kuhn and A. W. Tucker (Eds.), *Linear Inequalities and Related Systems* (Vol. 38 of *Annals of Mathematics Study*), pages 215–221. Princeton University Press, Princeton, NJ, 1956.

[157] J. N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *Ann. Math. Stat.*, pages 1470–1480, 1972.

[158] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

[159] L. D. Davisson. Universal noiseless coding. *IEEE Trans. Inf. Theory*, IT-19:783–795, 1973.

[160] L. D. Davisson. Minimax noiseless universal coding for Markov sources. *IEEE Trans. Inf. Theory*, pages 211–215, Mar. 1983.

[161] L. D. Davisson, R. J. McEliece, M. B. Pursley, and M. S. Wallace. Efficient universal noiseless source codes. *IEEE Trans. Inf. Theory*, pages 269–279, May 1981.

[162] A. Dembo. *Information Inequalities and Uncertainty Principles* (Technical Report), Department of Statistics, Stanford University, Stanford, CA, 1990.

[163] A. Dembo. Information inequalities and concentration of measure. *Ann. Prob.*, pages 927–939, 1997.

[164] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. Inf. Theory*, 37(6):1501–1518, Nov. 1991.

[165] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Jones & Bartlett, Boston, 1993.

[166] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39(1):1–38, 1977.

[167] L. Devroye and L. Gyorfi. *Nonparametric Density Estimation: The L_1 View.* Wiley, New York, 1985.

[168] L. Devroye, L. Gyorfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer-Verlag, New York, 1996.

[169] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum-channel capacity of very noisy channels. *Phys. Rev. A*, pages 830–839, 1998.

[170] R.L. Dobrushin. General formulation of Shannon's main theorem of information theory. *Usp. Math. Nauk*, 14:3–104, 1959. Translated in *Am. Math. Soc. Trans.*, 33:323-438.

[171] R. L. Dobrushin. Survey of Soviet research in information theory. *IEEE Trans. Inf. Theory*, pages 703–724, Nov. 1972.

[172] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*, pages 613–627, May 1995.

[173] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* Wiley, New York, 1973.

[174] G. Dueck. Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. *Probl. Contr. Inf. Theory*, pages 11–19, 1978.

[175] G. Dueck. The capacity region of the two-way channel can exceed the inner bound. *Inf. Control*, 40:258–266, 1979.

[176] G. Dueck. Partial feedback for two-way and broadcast channels. *Inf. Control*, 46:1–15, 1980.

[177] G. Dueck and J. Körner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Trans. Inf. Theory*, IT-25:82–85, 1979.

[178] P. M. Ebert. The capacity of the Gaussian channel with feedback. *Bell Syst. Tech. J.*, 49:1705–1712, Oct. 1970.

[179] P. M. Ebert. The capacity of the Gaussian channel with feedback. *Bell Syst. Tech. J.*, pages 1705–1712, Oct. 1970.

[180] K. Eckschlager. *Information Theory in Analytical Chemistry*. Wiley, New York, 1994.

[181] M. Effros, K. Visweswariah, S. R. Kulkarni, and S. Verdu. Universal lossless source coding with the Burrows–Wheeler transform. *IEEE Trans. Inf. Theory*, IT-48:1061–1081, May 2002.

[182] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, London, 1993.

[183] H. G. Eggleston. *Convexity* (Cambridge Tracts in Mathematics and Mathematical Physics, No. 47). Cambridge University Press, Cambridge, 1969.

[184] A. El Gamal. The feedback capacity of degraded broadcast channels. *IEEE Trans. Inf. Theory*, IT-24:379–381, 1978.

[185] A. El Gamal. The capacity region of a class of broadcast channels. *IEEE Trans. Inf. Theory*, IT-25:166–169, 1979.

[186] A. El Gamal and T. M. Cover. Multiple user information theory. *Proc. IEEE*, 68:1466–1483, 1980.

[187] A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory*, IT-28:851–857, 1982.

[188] A. El Gamal and E. C. Van der Meulen. A proof of Marton's coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, IT-27:120–122, 1981.

[189] P. Elias. Error-free coding. *IRE Trans. Inf. Theory*, IT-4:29–37, 1954.

[190] P. Elias. Coding for noisy channels. *IRE Conv. Rec., Pt. 4*, pages 37–46, 1955.

[191] P. Elias. Networks of Gaussian channels with applications to feedback systems. *IEEE Trans. Inf. Theory*, pages 493–501, July 1967.

[192] P. Elias. The efficient construction of an unbiased random sequence. *Ann. Math. Stat.*, pages 865–870, 1972.

[193] P. Elias. Universal codeword sets and representations of the integers. *IEEE Trans. Inf. Theory*, pages 194–203, Mar. 1975.

[194] P. Elias. Interval and recency rank source coding: two on-line adaptive variable-length schemes. *IEEE Trans. Inf. Theory*, pages 3–10, Jan. 1987.

[195] P. Elias, A. Feinstein, and C. E. Shannon. A note on the maximum flow through a network. *IEEE Trans. Inf. Theory*, pages 117–119, December 1956.

[196] R. S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer-Verlag, New York, 1985.

[197] A. Ephremides and B. Hajek. Information theory and communication networks: an unconsummated union. *IEEE Trans. Inf. Theory*, pages 2416–2434, Oct. 1998.

[198] W. H. R. Equitz and T. M. Cover. Successive refinement of information. *IEEE Trans. Inf. Theory*, pages 269–275, Mar. 1991.

[199] Ky Fan. On a theorem of Weyl concerning the eigenvalues of linear transformations II. *Proc. Nat. Acad. Sci. USA*, 36:31–35, 1950.

[200] Ky Fan. Some inequalities concerning positive-definite matrices. *Proc. Cambridge Philos. Soc.*, 51:414–421, 1955.

[201] R. M. Fano. Class notes for Transmission of Information, course 6.574 (Technical Report). MIT, Cambridge, MA, 1952.

[202] R. M. Fano. *Transmission of Information: A Statistical Theory of Communication*. Wiley, New York, 1961.

[203] M. Feder. A note on the competetive optimality of Huffman codes. *IEEE Trans. Inf. Theory*, 38(2):436–439, Mar. 1992.

[204] M. Feder, N. Merhav, and M. Gutman. Universal prediction of individual sequences. *IEEE Trans. Inf. Theory*, pages 1258–1270, July 1992.

[205] A. Feinstein. A new basic theorem of information theory. *IRE Trans. Inf. Theory*, IT-4:2–22, 1954.

[206] A. Feinstein. *Foundations of Information Theory*. McGraw-Hill, New York, 1958.

[207] A. Feinstein. On the coding theorem and its converse for finite-memory channels. *Inf. Control*, 2:25–44, 1959.

[208] W. Feller. *An Introduction to Probability Theory and Its Applications*, 2nd ed., Vol. 1. Wiley, New York, 1957.

[209] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc., London A*, 222:309–368, 1922.

[210] R. A. Fisher. Theory of statistical estimation. *Proc. Cambridge Philos. Soc.*, 22:700–725, 1925.

[211] B. M. Fitingof. Optimal encoding with unknown and variable message statistics. *Probl. Inf. Transm.* (*USSR*), pages 3–11, 1966.

[212] B. M. Fitingof. The compression of discrete information. *Probl. Inf. Transm.* (*USSR*), pages 28–36, 1967.

[213] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Can. J. Math.*, pages 399–404, 1956.

[214] L. R. Ford and D. R. Fulkerson. *Flows in Networks*. Princeton University Press, Princeton, NJ, 1962.

[215] G. D. Forney. Exponential error bounds for erasure, list and decision feedback schemes. *IEEE Trans. Inf. Theory*, IT-14:549–557, 1968.

[216] G. D. Forney. Information Theory: unpublished course notes. Stanford University, Stanford, CA, 1972.

[217] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Syst. Tech. J.*, 1(2):41–59, 1996.

[218] P. Franaszek, P. Tsoucas, and J. Thomas. Context allocation for multiple dictionary data compression. In *Proc. IEEE Int. Symp. Inf. Theory*, Trondheim, Norway, page 12, 1994.

[219] P. A. Franaszek. On synchronous variable length coding for discrete noiseless channels. *Inf. Control*, 15:155–164, 1969.

[220] T. Gaarder and J. K. Wolf. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE Trans. Inf. Theory*, IT-21:100–102, 1975.

[221] D. Gabor. Theory of communication. *J. Inst. Elec. Engg.*, pages 429–457, Sept. 1946.

[222] P. Gacs and J. Körner. Common information is much less than mutual information. *Probl. Contr. Inf. Theory*, pages 149–162, 1973.

[223] R. G. Gallager. Source coding with side information and universal coding. Unpublished manuscript, also presented at the Int. Symp. Inf. Theory, Oct. 1974.

[224] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Trans. Inf. Theory*, IT-11:3–18, 1965.

[225] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Probl. Peredachi Inf.*, 10(3):3–14, 1974.

[226] R. G. Gallager. Basic limits on protocol information in data communication networks. *IEEE Trans. Inf. Theory*, pages 385–398, July 1976.

[227] R. G. Gallager. A minimum delay routing algorithm using distributed computation. *IEEE Trans. Commun.*, pages 73–85, Jan. 1977.

[228] R. G. Gallager. Variations on a theme by Huffman. *IEEE Trans. Inf. Theory*, pages 668–674, Nov. 1978.

[229] R. G. Gallager. *Source Coding with Side Information and Universal Coding* (Tech. Rept. LIDS-P-937). Laboratory for Information Decision Systems, MIT, Cambridge, MA, 1979.

[230] R. G. Gallager. A perspective on multiaccess channels. *IEEE Trans. Inf. Theory*, pages 124–142, Mar. 1985.

[231] R. G. Gallager. Low density parity check codes. *IRE Trans. Inf. Theory*, IT-8:21–28, Jan. 1962.

[232] R. G. Gallager. *Low Density Parity Check Codes*. MIT Press, Cambridge, MA, 1963.

[233] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[234] A. A. El Gamal and T. M. Cover. Achievable rates for multiple descriptions. *IEEE Trans. Inf. Theory*, pages 851–857, November 1982.

[235] A. El Gamal. Broadcast channels with and without feedback. *11th Ann. Asilomar Conf. Circuits*, pages 180–183, Nov. 1977.

[236] A. El Gamal. Capacity of the product and sum of two unmatched broadcast channels. *Probl. Peredachi Inf.*, pages 3–23, Jan.–Mar. 1980.

[237] A. A. El Gamal. The feedback capacity of degraded broadcast channels (corresp.). *IEEE Trans. Inf. Theory*, pages 379–381, May 1978.

[238] A. A. El Gamal. The capacity of a class of broadcast channels. *IEEE Trans. Inf. Theory*, pages 166–169, Mar. 1979.

[239] A. A. El Gamal. The capacity of the physically degraded Gaussian broadcast channel with feedback (corresp.). *IEEE Trans. Inf. Theory*, pages 508–511, July 1981.

[240] A. A. El Gamal and E. C. van der Meulen. A proof of Marton's coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, pages 120–122, Jan. 1981.

[241] I. M. Gelfand, A. N. Kolmogorov, and A. M. Yaglom. On the general definition of mutual information. *Rept. Acad. Sci. USSR*, pages 745–748, 1956.

[242] S. I. Gelfand. Capacity of one broadcast channel. *Probl. Peredachi Inf.*, pages 106–108, July–Sept. 1977.

[243] S. I. Gelfand and M. S. Pinsker. Capacity of a broadcast channel with one deterministic component. *Probl. Peredachi Inf.*, pages 24–34, Jan.–Mar. 1980.

[244] S. I. Gelfand and M. S. Pinsker. Coding for channel with random parameters. *Probl. Contr. Inf. Theory*, pages 19–31, 1980.

[245] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression.* Kluwer, Boston, 1992.

[246] G. G. Rayleigh and J. M. Cioffi. Spatio-temporal coding for wireless communication. *IEEE Trans. Commun.*, 46:357–366, 1998.

[247] J. D. Gibson and J. L. Melsa. *Introduction to Nonparametric Detection with Applications*. IEEE Press, New York, 1996.

[248] E. N. Gilbert. Codes based on inaccurate source probabilities. *IEEE Trans. Inf. Theory*, pages 304–314, May 1971.

[249] E. N. Gilbert and E. F. Moore. Variable length binary encodings. *Bell Syst. Tech. J.*, 38:933–967, 1959.

[250] S. Goldman. Some fundamental considerations concerning noise reduction and range in radar and communication. *Proc. Inst. Elec. Engg.*, pages 584–594, 1948.

[251] S. Goldman. *Information Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1953.

[252] A. Goldsmith and M. Effros. The capacity region of Gaussian broadcast channels with intersymbol interference and colored Gaussian noise. *IEEE Trans. Inf. Theory*, 47:2–8, Jan. 2001.

[253] S. W. Golomb. Run-length encodings. *IEEE Trans. Inf. Theory*, pages 399–401, July 1966.

[254] S. W. Golomb, R. E. Peile, and R. A. Scholtz. *Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111* (*Applications of Communications Theory*). Plenum Publishing, New York, 1994.

[255] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting. Rate-splitting multiple access for discrete memoryless channels. *IEEE Trans. Inf. Theory*, pages 873–890, Mar. 2001.

[256] R. M. Gray. *Source Coding Theory.* Kluwer, Boston, 1990.

[257] R. M. Gray and L. D. Davisson, (Eds.). *Ergodic and Information Theory*. Dowden, Hutchinson & Ross, Stroudsburg, PA, 1977.

[258] R. M. Gray and Lee D. Davisson. Source coding theorems without the ergodic assumption. *IEEE Trans. Inf. Theory*, pages 502–516, July 1974.

[259] R. M. Gray. Sliding block source coding. *IEEE Trans. Inf. Theory*, IT-21:357–368, 1975.

[260] R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, New York, 1990.

[261] R. M. Gray and A. Wyner. Source coding for a simple network. *Bell Syst. Tech. J.*, 58:1681–1721, 1974.

[262] U. Grenander and G. Szego. *Toeplitz Forms and Their Applications*. University of California Press, Berkeley, CA, 1958.

[263] B. Grünbaum. *Convex Polytopes*. Interscience, New York, 1967.

[264] S. Guiasu. *Information Theory with Applications*. McGraw-Hill, New York, 1976.

[265] B. E. Hajek and M. B. Pursley. Evaluation of an achievable rate region for the broadcast channel. *IEEE Trans. Inf. Theory*, pages 36–46, Jan. 1979.

[266] R. V. Hamming. Error detecting and error correcting codes. *Bell Syst. Tech. J.*, 29:147–160, 1950.

[267] T. S. Han. The capacity region for the deterministic broadcast channel with a common message (corresp.). *IEEE Trans. Inf. Theory*, pages 122–125, Jan. 1981.

[268] T. S. Han and S. I. Amari. Statistical inference under multiterminal data compression. *IEEE Trans. Inf. Theory*, pages 2300–2324, Oct. 1998.

[269] T. S. Han and S. Verdu. New results in the theory of identification via channels. *IEEE Trans. Inf. Theory*, pages 14–25, Jan. 1992.

[270] T. S. Han. Nonnegative entropy measures of multivariate symmetric correlations. *Inf. Control*, 36(2):133–156, 1978.

[271] T. S. Han. The capacity region of a general multiple access channel with certain correlated sources. *Inf. Control*, 40:37–60, 1979.

[272] T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer-Verlag, New York, 2002.

[273] T. S. Han and M. H. M. Costa. Broadcast channels with arbitrarily correlated sources. *IEEE Trans. Inf. Theory*, IT-33:641–650, 1987.

[274] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theory*, IT-27:49–60, 1981.

[275] R. V. Hartley. Transmission of information. *Bell Syst. Tech. J.*, 7:535, 1928.

[276] C. W. Helstrom. *Elements of Signal Detection and Estimation*. Prentice-Hall, Englewood Cliffs, NJ, 1995.

[277] Y. Hershkovits and J. Ziv. On sliding-window universal data compression with limited memory. *IEEE Trans. Inf. Theory*, pages 66–78, Jan. 1998.

[278] P. A. Hocquenghem. Codes correcteurs d'erreurs. *Chiffres*, 2:147–156, 1959.

[279] J. L. Holsinger. *Digital Communication over Fixed Time-Continuous Channels with Memory, with Special Application to Telephone Channels* (Technical Report). MIT, Cambridge, MA, 1964.

[280] M. L. Honig, U. Madhow, and S. Verdu. Blind adaptive multiuser detection. *IEEE Trans. Inf. Theory*, pages 944–960, July 1995.

[281] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Formal Languages and Computation*. Addison-Wesley, Reading, MA, 1979.

[282] Y. Horibe. An improved bound for weight-balanced tree. *Inf. Control*, 34:148–151, 1977.

[283] D. A. Huffman. A method for the construction of minimum redundancy codes. *Proc. IRE*, 40:1098–1101, 1952.

[284] J. Y. Hui. *Switching an Traffic Theory for Integrated Broadband Networks.* Kluwer, Boston, 1990.

[285] J. Y. N. Hui and P. A. Humblet. The capacity region of the totally asynchronous multiple-access channel. *IEEE Trans. Inf. Theory*, pages 207–216, Mar. 1985.

[286] S. Ihara. On the capacity of channels with additive non-Gaussian noise. *Inf. Contr.*, pages 34–39, 1978.

[287] S. Ihara. *Information Theory for Continuous Systems*. World Scientific, Singapore, 1993.

[288] K. A. Schouhamer Immink, Paul H. Siegel, and Jack K. Wolf. Codes for digital recorders. *IEEE Trans. Inf. Theory*, pages 2260–2299, Oct. 1998.

[289] N. S. Jayant (Ed.). *Waveform Quantization and Coding*. IEEE Press, New York, 1976.

[290] N. S. Jayant and P. Noll. *Digital Coding of Waveforms*. Prentice-Hall, Englewood Cliffs, NJ, 1984.

[291] E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620, 1957.

[292] E. T. Jaynes. Information theory and statistical mechanics II. *Phys. Rev.*, 108:171, 1957.

[293] E. T. Jaynes. On the rationale of maximum entropy methods. *Proc. IEEE*, 70:939–952, 1982.

[294] E. T. Jaynes. *Papers on Probability, Statistics and Statistical Physics*. Reidel, Dordrecht, The Netherlands, 1982.

[295] F. Jelinek. Buffer overflow in variable length encoding of fixed rate sources. *IEEE Trans. Inf. Theory*, IT-14:490–501, 1968.

[296] F. Jelinek. Evaluation of expurgated error bounds. *IEEE Trans. Inf. Theory*, IT-14:501–505, 1968.

[297] F. Jelinek. *Probabilistic Information Theory*. McGraw-Hill, New York, 1968.

[298] F. Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, MA, 1998.

[299] R Jozsa and B. Schumacher. A new proof of the quantum noiseless coding theorem. *J Mod. Opt.*, pages 2343–2350, 1994.

[300] G. G. Langdon, Jr. A note on the Ziv–Lempel model for compressing individual sequences. *IEEE Trans. Inf. Theory*, pages 284–287, Mar. 1983.

[301] J. Justesen. A class of constructive asymptotically good algebraic codes. *IEEE Trans. Inf. Theory*, IT-18:652–656, 1972.

[302] M. Kac. On the notion of recurrence in discrete stochastic processes. *Bull. Am. Math. Soc.*, pages 1002–1010, Oct. 1947.

[303] T. Kailath and J. P. M. Schwalkwijk. A coding scheme for additive noise channels with feedback. Part I: No bandwidth constraints. *IEEE Trans. Inf. Theory*, IT-12:172–182, 1966.

[304] T. Kailath and H. V. Poor. Detection of stochastic processes. *IEEE Trans. Inf. Theory*, pages 2230–2259, Oct. 1998.

[305] S. Karlin. *Mathematical Methods and Theory in Games, Programming and Economics*, Vol. 2. Addison-Wesley, Reading, MA, 1959.

[306] J. Karush. A simple proof of an inequality of McMillan. *IRE Trans. Inf. Theory*, IT-7:118, 1961.

[307] F. P. Kelly. Notes on effective bandwidth. *Stochastic Networks Theory and Applications*, pages 141–168, 1996.

[308] J. Kelly. A new interpretation of information rate. *Bell Syst. Tech. J*, 35:917–926, July 1956.

[309] J. H. B. Kemperman. *On the Optimum Rate of Transmitting Information* (Lecture Notes in Mathematics), pages 126–169. Springer Verlag, New York, 1967.

[310] M. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Macmillan, New York, 1977.

[311] A. Y. Khinchin. *Mathematical Foundations of Information Theory*. Dover, New York, 1957.

[312] J. C. Kieffer. A simple proof of the Moy–Perez generalization of the Shannon–McMillan theorem. *Pacific J. Math.*, 51:203–206, 1974.

[313] J. C. Kieffer. A survey of the theory of source coding with a fidelity criterion. *IEEE Trans. Inf. Theory*, pages 1473–1490, Sept. 1993.

[314] Y. H. Kim. Feedback capacity of first-order moving average Gaussian channel. *Proc. IEEE Int. Symp. Information Theory*, Adelaide, pages 416–420, Sept. 2005.

[315] D. E. Knuth. Dynamic Huffman coding. *J. Algorithms*, pages 163–180, 1985.

[316] D. E. Knuth. *Art of Computer Programming*.

[317] D. E. Knuth and A. C. Yao. The complexity of random number generation. In J. F. Traub (Ed.), *Algorithms and Complexity: Recent Results and New Directions* (Proceedings of the Symposium on New Directions and Recent Results in Algorithms and Complexity, Carnegie-Mellon University, 1976), pages 357–428. Academic Press, New York, 1976.

[318] A. N. Kolmogorov. A new metric invariant of transitive dynamical systems and automorphism in Lebesgue spaces. *Dokl. Akad. Nauk SSSR*, pages 861–864, 1958.

[319] A. N. Kolmogorov. On the Shannon theory of information transmission in the case of continuous signals. *IRE Trans. Inf. Theory*, IT-2:102–108, Sept. 1956.

[320] A. N. Kolmogorov. A new invariant for transitive dynamical systems. *Dokl. Acad. Nauks SSR*, 119:861–864, 1958.

[321] A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Probl. Inf. Transm. (USSR)*, 1:4–7, 1965.

[322] A. N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Trans. Inf. Theory*, IT-14:662–664, 1968.

[323] A. N. Kolmogorov. The theory of transmission of information. In *Selected Works of A. N. Kolmogorov, Vol. III: Information Theory and the Theory of Algorithms*, Session on scientific problems of automatization in industry, Vol. 1, Plenary talks, Izd. Akad. Nauk SSSR, Moscow, 1957, pages 66–99. Kluwer, Dordrecht, The Netherlands, 1993.

[324] J. Körner and K. Marton. The comparison of two noisy channels. In I. Csiszár and P. Elias (Ed.), *Topics in Information Theory* (Coll. Math. Soc. J. Bolyai, No. 16), pages 411–423. North-Holland, Amsterdam, 1977.

[325] J. Körner and K. Marton. General broadcast channels with degraded message sets. *IEEE Trans. Inf. Theory*, IT-23:60–64, 1977.

[326] J. Körner and K. Marton. How to encode the modulo 2 sum of two binary sources. *IEEE Trans. Inf. Theory*, IT-25:219–221, 1979.

[327] J. Körner and A. Orlitsky. Zero error information theory. *IEEE Trans. Inf. Theory*, IT-44:2207–2229, Oct. 1998.

[328] V. A. Kotel'nikov. On the transmission capacity of "ether" and wire in electrocommunications. *Izd. Red. Upr. Svyazi RKKA*, 44, 1933.

[329] V. A. Kotel'nikov. *The Theory of Optimum Noise Immunity*. McGraw-Hill, New York, 1959.

[330] L. G. Kraft. *A device for quantizing, grouping and coding amplitude modulated pulses*. Master's thesis, Department of Electrical Engineering, MIT, Cambridge, MA, 1949.

[331] R. E. Krichevsky. Laplace's law of succession and universal encoding. *IEEE Trans. Inf. Theory*, pages 296–303, Jan. 1998.

[332] R. E. Krichevsky. *Universal Compression and Retrieval*. Kluwer, Dordrecht, The Netherlands, 1994.

[333] R. E. Krichevsky and V. K. Trofimov. The performance of universal encoding. *IEEE Trans. Inf. Theory*, pages 199–207, Mar. 1981.

[334] S. R. Kulkarni, G. Lugosi, and S. S. Venkatesh. Learning pattern classification: a survey. *IEEE Trans. Inf. Theory*, pages 2178–2206, Oct. 1998.

[335] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.

[336] S. Kullback. A lower bound for discrimination in terms of variation. *IEEE Trans. Inf. Theory*, IT-13:126–127, 1967.

[337] S. Kullback, J. C. Keegel, and J. H. Kullback. *Topics in Statistical Information Theory*. Springer-Verlag, Berlin, 1987.

[338] S. Kullback and M. A. Khairat. A note on minimum discrimination information. *Ann. Math. Stat.*, pages 279–280, 1966.

[339] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[340] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part II. *Bell Syst. Tech. J.*, 40:65–84, 1961.

[341] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part III. *Bell Syst. Tech. J.*, 41:1295–1336, 1962.

[342] G. G. Langdon. An introduction to arithmetic coding. *IBM J. Res. Dev.*, 28:135–149, 1984.

[343] G. G. Langdon and J. J. Rissanen. A simple general binary source code. *IEEE Trans. Inf. Theory*, IT-28:800, 1982.

[344] A. Lapidoth and P. Narayan. Reliable communication under channel uncertainty. *IEEE Trans. Inf. Theory*, pages 2148–2177, Oct. 1998.

[345] A. Lapidoth and J. Ziv. On the universality of the LZ-based decoding algorithm. *IEEE Trans. Inf. Theory*, pages 1746–1755, Sept. 1998.

[346] H. A. Latané. Criteria for choice among risky ventures. *J. Polit. Econ.*, 38:145–155, Apr. 1959.

[347] H. A. Latané and D.L. Tuttle. Criteria for portfolio building. *J. Finance*, 22:359–373, Sept. 1967.

[348] E. A. Lee and D. G. Messerschmitt. *Digital Communication*, 2nd ed. Kluwer, Boston, 1994.

[349] J. Leech and N. J. A. Sloane. Sphere packing and error-correcting codes. *Can. J. Math*, pages 718–745, 1971.

[350] E. L. Lehmann and H. Scheffé. Completeness, similar regions and unbiased estimation. *Sankhya*, 10:305–340, 1950.

[351] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Trans. Inf. Theory*, pages 75–81, Jan. 1976.

[352] L. A. Levin. On the notion of a random sequence. *Sov. Math. Dokl.*, 14:1413–1416, 1973.

[353] L. A. Levin and A. K. Zvonkin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russ. Math. Surv.*, 25/6:83–124, 1970.

[354] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd ed. Springer-Verlag, New York, 1997.

[355] H. Liao. *Multiple access channels*. Ph.D. thesis, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[356] S. Lin and D. J. Costello, Jr. *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, Englewood Cliffs, NJ, 1983.

[357] D. Lind and B. Marcus. *Symbolic Dynamics and Coding*. Cambridge University Press, Cambridge, 1995.

[358] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Trans. Commun.*, COM-28:84–95, 1980.

[359] T. Linder, G. Lugosi, and K. Zeger. Rates of convergence in the source coding theorem in empirical quantizer design. *IEEE Trans. Inf. Theory*, pages 1728–1740, Nov. 1994.

[360] T. Linder, G. Lugosi, and K. Zeger. Fixed-rate universal lossy source coding and rates of convergence for memoryless sources. *IEEE Trans. Inf. Theory*, pages 665–676, May 1995.

[361] D. Lindley. *Boltzmann's Atom: The Great Debate That Launched A Revolution in Physics*. Free Press, New York, 2001.

[362] A. Liversidge. Profile of Claude Shannon. In N. J. A. Sloane and A. D. Wyner (Eds.), *Claude Elwood Shannon Collected Papers*. IEEE Press, Piscataway, NJ, 1993 (*Omni* magazine, Aug. 1987.)

[363] S. P. Lloyd. Least Squares Quantization in PCM (Technical Report). *Bell Lab. Tech. Note*, 1957.

[364] G. Louchard and Wojciech Szpankowski. On the average redundancy rate of the Lempel–Ziv code. *IEEE Trans. Inf. Theory*, pages 2–8, Jan. 1997.

[365] L. Lovasz. On the Shannon capacity of a graph. *IEEE Trans. Inf. Theory*, IT-25:1–7, 1979.

[366] R. W. Lucky. *Silicon Dreams: Information, Man and Machine*. St. Martin's Press, New York, 1989.

[367] D. J. C. Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, 2003.

[368] D. J. C. MacKay and R. M. Neal. Near Shannon limit performance of low-density parity-check codes. *Electron. Lett.*, pages 1645–1646, Mar. 1997.

[369] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, Amsterdam, 1977.

[370] B. Marcus. Sofic systems and encoding data. *IEEE Trans. Inf. Theory*, IT-31(3):366–377, May 1985.

[371] R. J. Marks. *Introduction to Shannon Sampling and Interpolation Theory*. Springer-Verlag New York, 1991.

[372] A. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York, 1979.

[373] A. Marshall and I. Olkin. A convexity proof of Hadamard's inequality. *Am. Math. Monthly*, 89(9):687–688, 1982.

[374] P. Martin-Löf. The definition of random sequences. *Inf. Control*, 9:602–619, 1966.

[375] K. Marton. Information and information stability of ergodic sources. *Probl. Inf. Transm.* (*VSSR*), pages 179–183, 1972.

[376] K. Marton. Error exponent for source coding with a fidelity criterion. *IEEE Trans. Inf. Theory*, IT-20:197–199, 1974.

[377] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, IT-25:306–311, 1979.

[378] J. L. Massey and P. Mathys. The collision channel without feedback. *IEEE Trans. Inf. Theory*, pages 192–204, Mar. 1985.

[379] R. A. McDonald. *Information rates of Gaussian signals under criteria constraining the error spectrum*. D. Eng. dissertation, Yale University School of Electrical Engineering, New Haven, CT, 1961.

[380] R. A. McDonald and P. M. Schultheiss. Information rates of Gaussian signals under criteria constraining the error spectrum. *Proc. IEEE*, pages 415–416, 1964.

[381] R. A. McDonald and P. M. Schultheiss. Information rates of Gaussian signals under criteria constraining the error spectrum. *Proc. IEEE*, 52:415–416, 1964.

[382] R. J. McEliece, D. J. C. MacKay, and J. F. Cheng. Turbo decoding as an instance of Pearl's belief propagation algorithm. *IEEE J. Sel. Areas Commun.*, pages 140–152, Feb. 1998.

[383] R. J. McEliece. *The Theory of Information and Coding*. Addison-Wesley, Reading, MA, 1977.

[384] B. McMillan. The basic theorems of information theory. *Ann. Math. Stat.*, 24:196–219, 1953.

[385] B. McMillan. Two inequalities implied by unique decipherability. *IEEE Trans. Inf. Theory*, IT-2:115–116, 1956.

[386] N. Merhav and M. Feder. Universal schemes for sequential decision from individual data sequences. *IEEE Trans. Inf. Theory*, pages 1280–1292, July 1993.

[387] N. Merhav and M. Feder. A strong version of the redundancy-capacity theorem of universal coding. *IEEE Trans. Inf. Theory*, pages 714–722, May 1995.

[388] N. Merhav and M. Feder. Universal prediction. *IEEE Trans. Inf. Theory*, pages 2124–2147, Oct. 1998.

[389] R. C. Merton and P. A. Samuelson. Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods. *J. Finan. Econ.*, 1:67–94, 1974.

[390] H. Minkowski. Diskontinuitätsbereich für arithmetische Äquivalenz. *J. Math.*, 129:220–274, 1950.

[391] L. Mirsky. On a generalization of Hadamard's determinantal inequality due to Szasz. *Arch. Math.*, VIII:274–275, 1957.

[392] S. C. Moy. Generalizations of the Shannon–McMillan theorem. *Pacific J. Math.*, pages 705–714, 1961.

[393] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, NJ, 1980.

[394] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. Roy. Soc. London A*, 231:289–337, 1933.

[395] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[396] H. Nyquist. Certain factors affecting telegraph speed. *Bell Syst. Tech. J.*, 3:324, 1924.

[397] H. Nyquist. Certain topics in telegraph transmission theory. *AIEE Trans.*, pages 617–644, Apr. 1928.

[398] J. Omura. A coding theorem for discrete time sources. *IEEE Trans. Inf. Theory*, IT-19:490–498, 1973.

[399] A. Oppenheim. Inequalities connected with definite Hermitian forms. *J. London Math. Soc.*, 5:114–119, 1930.

[400] E. Ordentlich. On the factor-of-two bound for Gaussian multiple-access channels with feedback. *IEEE Trans. Inf. Theory*, pages 2231–2235, Nov. 1996.

[401] E. Ordentlich and T. Cover. The cost of achieving the best portfolio in hindsight. *Math. Operations Res.*, 23(4): 960–982, Nov. 1998.

[402] S. Orey. On the Shannon–Perez–Moy theorem. *Contemp. Math.*, 41:319–327, 1985.

[403] A. Orlitsky. Worst-case interactive communication. I: Two messages are almost optimal. *IEEE Trans. Inf. Theory*, pages 1111–1126, Sept. 1990.

[404] A. Orlitsky. Worst-case interactive communication. II: Two messages are not optimal. *IEEE Trans. Inf. Theory*, pages 995–1005, July 1991.

[405] A. Orlitsky. Average-case interactive communication. *IEEE Trans. Inf. Theory*, pages 1534–1547, Sept. 1992.

[406] A. Orlitsky and A. El Gamal. Average and randomized communication complexity. *IEEE Trans. Inf. Theory*, pages 3–16, Jan. 1990.

[407] D. S. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.*, pages 337–352, 1970.

[408] D. S. Ornstein and B. Weiss. Entropy and data compression schemes. *IEEE Trans. Inf. Theory*, pages 78–83, Jan. 1993.

[409] D. S. Ornstein. Bernoulli shifts with the same entropy are isomorphic. *Adv. Math.*, 4:337–352, 1970.

[410] L. H. Ozarow. The capacity of the white Gaussian multiple access channel with feedback. *IEEE Trans. Inf. Theory*, IT-30:623–629, 1984.

[411] L. H. Ozarow and C. S. K. Leung. An achievable region and an outer bound for the Gaussian broadcast channel with feedback. *IEEE Trans. Inf. Theory*, IT-30:667–671, 1984.

[412] H. Pagels. *The Dreams of Reason: the Computer and the Rise of the Sciences of Complexity*. Simon and Schuster, New York, 1988.

[413] C. Papadimitriou. *Information theory and computational complexity: The expanding interface*. *IEEE Inf. Theory Newslett.* (Special Golden Jubilee Issue), pages 12–13, June 1998.

[414] R. Pasco. *Source coding algorithms for fast data compression*. Ph.D. thesis, Stanford University, Stanford, CA, 1976.

[415] A. J. Paulraj and C. B. Papadias. Space-time processing for wireless communications. *IEEE Signal Processing Mag.*, pages 49–83, Nov. 1997.

[416] W. B. Pennebaker and J. L. Mitchell. *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York, 1988.

[417] A. Perez. Extensions of Shannon–McMillan's limit theorem to more general stochastic processes. In *Trans. Third Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 545–574, Czechoslovak Academy of Sciences, Prague, 1964.

[418] J. R. Pierce. The early days of information theory. *IEEE Trans. Inf. Theory*, pages 3–8, Jan. 1973.

[419] J. R. Pierce. *An Introduction to Information Theory: Symbols, Signals and Noise*, 2nd ed. Dover Publications, New York, 1980.

[420] J. T. Pinkston. An application of rate-distortion theory to a converse to the coding theorem. *IEEE Trans. Inf. Theory*, IT-15:66–71, 1969.

[421] M. S. Pinsker. Talk at Soviet Information Theory meeting, 1969. No abstract published.

[422] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, CA, 1964. (Originally published in Russian in 1960.)

[423] M. S. Pinsker. The capacity region of noiseless broadcast channels. *Probl. Inf. Transm. (USSR)*, 14(2):97–102, 1978.

[424] M. S. Pinsker and R. L. Dobrushin. Memory increases capacity. *Probl. Inf. Transm. (USSR)*, pages 94–95, Jan. 1969.

[425] M. S. Pinsker. *Information and Stability of Random Variables and Processes*. Izd. Akad. Nauk, 1960. Translated by A. Feinstein, 1964.

[426] E. Plotnik, M. Weinberger, and J. Ziv. Upper bounds on the probability of sequences emitted by finite-state sources and on the redundancy of the Lempel–Ziv algorithm. *IEEE Trans. Inf. Theory*, IT-38(1):66–72, Jan. 1992.

[427] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

[428] G. S. Poltyrev. Carrying capacity for parallel broadcast channels with degraded components. *Probl. Peredachi Inf.*, pages 23–35, Apr.–June 1977.

[429] S. Pombra and T. M. Cover. Nonwhite Gaussian multiple access channels with feedback. *IEEE Trans. Inf. Theory*, pages 885–892, May 1994.

[430] H. V. Poor. *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer-Verlag, New York, 1994.

[431] F. Pratt. *Secret and Urgent*. Blue Ribbon Books, Garden City, NY, 1939.

[432] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, pages 257–286, Feb. 1989.

[433] L. R. Rabiner and R. W. Schafer. *Digital Processing of Speech Signals*. Prentice-Hall, Englewood Cliffs, NJ, 1978.

[434] R. Ahlswede and Z. Zhang. New directions in the theory of identification via channels. *IEEE Trans. Inf. Theory*, 41:1040–1050, 1995.

[435] C. R. Rao. Information and accuracy obtainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[436] I. S. Reed. *1982 Claude Shannon lecture: Application of transforms to coding and related topics*. *IEEE Inf. Theory Newslett.*, pages 4–7, Dec. 1982.

[437] F. M. Reza. *An Introduction to Information Theory*. McGraw-Hill, New York, 1961.

[438] S. O. Rice. Mathematical analysis of random noise. *Bell Syst. Tech. J.*, pages 282–332, Jan. 1945.

[439] S. O. Rice. Communication in the presence of noise: probability of error for two encoding schemes. *Bell Syst. Tech. J.*, 29:60–93, 1950.

[440] B. E. Rimoldi and R. Urbanke. A rate-splitting approach to the Gaussian multiple-access channel. *IEEE Trans. Inf. Theory*, pages 364–375, Mar. 1996.

[441] J. Rissanen. Generalized Kraft inequality and arithmetic coding. *IBM J. Res. Dev.*, 20:198, 1976.

[442] J. Rissanen. Modelling by shortest data description. *Automatica*, 14:465–471, 1978.

[443] J. Rissanen. A universal prior for integers and estimation by minimum description length. *Ann. Stat.*, 11:416–431, 1983.

[444] J. Rissanen. Universal coding, information, prediction and estimation. *IEEE Trans. Inf. Theory*, IT-30:629–636, 1984.

[445] J. Rissanen. Stochastic complexity and modelling. *Ann. Stat.*, 14:1080–1100, 1986.

[446] J. Rissanen. Stochastic complexity (with discussions). *J. Roy. Stat. Soc.*, 49:223–239, 252–265, 1987.

[447] J. Rissanen. *Stochastic complexity in Statistical Inquiry*. World Scientific, Singapore, 1989.

[448] J. J. Rissanen. Complexity of strings in the class of Markov sources. *IEEE Trans. Inf. Theory*, pages 526–532, July 1986.

[449] J. J. Rissanen and G. G. Langdon, Jr. Universal modeling and coding. *IEEE Trans. Inf. Theory*, pages 12–23, Jan. 1981.

[450] B. Y. Ryabko. Encoding a source with unknown but ordered probabilities. *Probl. Inf. Transm.*, pages 134–139, Oct. 1979.

[451] B. Y. Ryabko. A fast on-line adaptive code. *IEEE Trans. Inf. Theory*, pages 1400–1404, July 1992.

[452] P. A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Rev. Econ. Stat.*, pages 236–239, 1969.

[453] P. A. Samuelson. The "fallacy" of maximizing the geometric mean in long sequences of investing or gambling. *Proc. Natl. Acad. Sci. USA*, 68:214–224, Oct. 1971.

[454] P. A. Samuelson. Why we should not make mean log of wealth big though years to act are long. *J. Banking and Finance*, 3:305–307, 1979.

[455] I. N. Sanov. On the probability of large deviations of random variables. *Mat. Sbornik*, 42:11–44, 1957. English translation in *Sel. Transl. Math. Stat. Prob.,* Vol. 1, pp. 213-244, 1961.

[456] A. A. Sardinas and G.W. Patterson. A necessary and sufficient condition for the unique decomposition of coded messages. *IRE Conv. Rec., Pt. 8*, pages 104–108, 1953.

[457] H. Sato. On the capacity region of a discrete two-user channel for strong interference. *IEEE Trans. Inf. Theory*, IT-24:377–379, 1978.

[458] H. Sato. The capacity of the Gaussian interference channel under strong interference. *IEEE Trans. Inf. Theory*, IT-27:786–788, 1981.

[459] H. Sato and M. Tanabe. A discrete two-user channel with strong interference. *Trans. IECE Jap.*, 61:880–884, 1978.

[460] S. A. Savari. Redundancy of the Lempel–Ziv incremental parsing rule. *IEEE Trans. Inf. Theory*, pages 9–21, January 1997.

[461] S. A. Savari and R. G. Gallager. Generalized Tunstall codes for sources with memory. *IEEE Trans. Inf. Theory*, pages 658–668, Mar. 1997.

[462] K. Sayood. *Introduction to Data Compression*. Morgan Kaufmann, San Francisco, CA, 1996.

[463] J. P. M. Schalkwijk. A coding scheme for additive noise channels with feedback. II: Bandlimited signals. *IEEE Trans. Inf. Theory*, pages 183–189, Apr. 1966.

[464] J. P. M. Schalkwijk. The binary multiplying channel: a coding scheme that operates beyond Shannon's inner bound. *IEEE Trans. Inf. Theory*, IT-28:107–110, 1982.

[465] J. P. M. Schalkwijk. On an extension of an achievable rate region for the binary multiplying channel. *IEEE Trans. Inf. Theory*, IT-29:445–448, 1983.

[466] C. P. Schnorr. A unified approach to the definition of random sequences. *Math. Syst. Theory*, 5:246–258, 1971.

[467] C. P. Schnorr. Process, complexity and effective random tests. *J. Comput. Syst. Sci.*, 7:376–388, 1973.

[468] C. P. Schnorr. A surview on the theory of random sequences. In R. Butts and J. Hinitikka (Eds.), *Logic, Methodology and Philosophy of Science*. Reidel, Dordrecht, The Netherlands, 1977.

[469] G. Schwarz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.

[470] S. Shamai and S. Verdu. The empirical distribution of good codes. *IEEE Trans. Inf. Theory*, pages 836–846, May 1997.

[471] C. E. Shannon. *A Mathematical Theory of Cryptography* (Tech. Rept. MM 45-110-02). Bell Lab. Tech. Memo., Sept. 1, 1945.

[472] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423,623–656, 1948.

[473] C. E. Shannon. Some geometrical results in channel capacity. *Verh. Dtsch. Elektrotechnik. Fachber.*, pages 13–15, 1956.

[474] C. E. Shannon. The zero-error capacity of a noisy channel. *IRE Trans. Inf. Theory*, IT-2:8–19, 1956.

[475] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Dev.*, pages 289–293, 1958.

[476] C. E. Shannon. Probability of error for optimal codes in a Gaussian channel. *Bell Syst. Tech. J.*, pages 611–656, May 1959.

[477] C. E. Shannon. Two-way communication channels. *Proc. 4th. Berkeley Symp. Mathematical Statistics and Probability* (June 20–July 30, 1960), pages 611–644, 1961.

[478] C. E. Shannon. *The wonderful world of feedback*. IEEE Int. Symp. Infor. Theory, ser. First Shannon Lecture, Ashkelon, Israel, 1973.

[479] C. E. Shannon. The mind reading machine. *In Shannon's Collected Papers*, pages 688–689, 1993.

[480] C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37:10–21, January 1949.

[481] C. E. Shannon. Communication theory of secrecy systems. *Bell Syst. Tech. J.*, 28:656–715, 1949.

[482] C. E. Shannon. Prediction and entropy of printed English. *Bell Syst. Tech. J.*, 30:50–64, January 1951.

[483] C. E. Shannon. Certain results in coding theory for noisy channels. *Infor. Control*, 1:6–25, 1957.

[484] C. E. Shannon. Channels with side information at the transmitter. *IBM J. Res. Dev.*, 2:289–293, 1958.

[485] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec., Pt. 4*, pages 142–163, 1959.

[486] C. E. Shannon. Two-way communication channels. In *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, Vol. 1, pages 611–644. University of California Press, Berkeley, CA, 1961.

[487] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding in discrete memoryless channels. I. *Inf. Control*, 10:65–103, 1967.

[488] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding in discrete memoryless channels. II. *Inf. Control*, 10:522–552, 1967.

[489] C. E. Shannon and W. W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, 1949.

[490] C. E. Shannon. General treatment of the problem of coding. *IEEE Trans. Inf. Theory*, pages 102–104, February 1953.

[491] W. F. Sharpe. *Investments*, 3rd ed. Prentice-Hall, Englewood Cliffs, NJ, 1985.

[492] P. C. Shields. Universal redundancy rates do not exist. *IEEE Trans. Inf. Theory*, pages 520–524, Mar. 1993.

[493] P. C. Shields. The interactions between ergodic theory and information theory. *IEEE Trans. Inf. Theory*, pages 2079–2093, Oct. 1998.

[494] P. C. Shields and B. Weiss. Universal redundancy rates for the class of B-processes do not exist. *IEEE Trans. Inf. Theory*, pages 508–512, Mar. 1995.

[495] J. E. Shore and R. W. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory*, IT-26:26–37, 1980.

[496] Y. M. Shtarkov. Universal sequential coding of single messages. *Probl. Inf. Transm.* (*USSR*), 23(3):3–17, July–Sept. 1987.

[497] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis, Queues, Communication and Computing*. Chapman & Hall, London, 1995.

[498] D. Slepian. *Key Papers in the Development of Information Theory*. IEEE Press, New York, 1974.

[499] D. Slepian. On bandwidth. *Proc. IEEE*, pages 292–300, Mar. 1976.

[500] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part I. *Bell Syst. Tech. J.*, 40:43–64, 1961.

[501] D. Slepian and J. K. Wolf. A coding theorem for multiple access channels with correlated sources. *Bell Syst. Tech. J.*, 52:1037–1076, 1973.

[502] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, IT-19:471–480, 1973.

[503] D. S. Slepian. Information theory in the fifties. *IEEE Trans. Inf. Theory*, pages 145–148, Mar. 1973.

[504] R. J. Solomonoff. A formal theory of inductive inference. *Inf. Control*, 7:1–22,224–254, 1964.

[505] A. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inf. Control*, 2:101–112, June 1959.

[506] A. Steane. Quantum computing. *Rept. Progr. Phys.*, pages 117–173, Feb. 1998.

[507] J. A. Storer and T. G. Szymanski. Data compression via textual substitution. *J. ACM*, 29(4):928–951, 1982.

[508] W. Szpankowski. Asymptotic properties of data compression and suffix trees. *IEEE Trans. Inf. Theory*, pages 1647–1659, Sept. 1993.

[509] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley-Interscience, New York, 2001.

[510] D. L. Tang and L. R. Bahl. Block codes for a class of constrained noiseless channels. *Inf. Control*, 17:436–461, 1970.

[511] I. E. Teletar and R. G. Gallager. Combining queueing theory with information theory for multiaccess. *IEEE J. Sel. Areas Commun.*, pages 963–969, Aug. 1995.

[512] E. Teletar. Capacity of multiple antenna Gaussian channels. *Eur. Trans. Telecommun.*, 10(6):585–595, 1999.

[513] J. A. Thomas. Feedback can at most double Gaussian multiple access channel capacity. *IEEE Trans. Inf. Theory*, pages 711–716, Sept. 1987.

[514] T. J. Tjalkens and F. M. J. Willems. A universal variable-to-fixed length source code based on Lawrence's algorithm. *IEEE Trans. Inf. Theory*, pages 247–253, Mar. 1992.

[515] T. J. Tjalkens and F. M. J. Willems. Variable- to fixed-length codes for Markov sources. *IEEE Trans. Inf. Theory*, pages 246–257, Mar. 1987.

[516] S. C. Tornay. *Ockham: Studies and Selections* (chapter "Commentarium in Sententias," I, 27). Open Court Publishers, La Salle, IL, 1938.

[517] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*, Part I. Wiley, New York, 1968.

[518] B. S. Tsybakov. Capacity of a discrete-time Gaussian channel with a filter. *Probl. Inf. Transm.*, pages 253–256, July–Sept. 1970.

[519] B. P. Tunstall. *Synthesis of noiseless compression codes*. Ph.D. dissertation, Georgia Institute of Technology, Atlanta, GA, Sept. 1967.

[520] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Trans. Inf. Theory*, pages 55–67, January 1982.

[521] G. Ungerboeck. Trellis-coded modulation with redundant signal sets part I: Introduction. *IEEE Commun. Mag.*, pages 5–11, Feb. 1987.

[522] G. Ungerboeck. Trellis-coded modulation with redundant signal sets part II: State of the art. *IEEE Commun. Mag.*, pages 12–21, Feb. 1987.

[523] I. Vajda. *Theory of Statistical Inference and Information*. Kluwer, Dordrecht, The Netherlands, 1989.

[524] L. G. Valiant. A theory of the learnable. *Commun. ACM*, pages 1134–1142, 1984.

[525] J. M. Van Campenhout and T. M. Cover. Maximum entropy and conditional probability. *IEEE Trans. Inf. Theory*, IT-27:483–489, 1981.

[526] E. Van der Meulen. Random coding theorems for the general discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, IT-21:180–190, 1975.

[527] E. C. van der Meulen. *Some reflections on the interference channel.* In R. E. Blahut, D. J. Costello, U. Maurer, and T. Mittelholzer, (Eds.), *Communications and Cryptography: Two Sides of One Tapestry*. Kluwer, Boston, 1994.

[528] E. C. Van der Meulen. A survey of multi-way channels in information theory. *IEEE Trans. Inf. Theory*, IT-23:1–37, 1977.

[529] E. C. Van der Meulen. Recent coding theorems for multi-way channels. Part I: The broadcast channel (1976-1980). In J. K. Skwyrzinsky (Ed.), *New Concepts in Multi-user Communication* (NATO Advanced Study Institute Series), pages 15–51. Sijthoff & Noordhoff, Amsterdam, 1981.

[530] E. C. Van der Meulen. Recent coding theorems and converses for multi-way channels. Part II: The multiple access channel (1976–1985) (Technical Report). Department Wiskunde, Katholieke Universiteit Leuven, 1985.

[531] V. N. Vapnik. *Estimation of Dependencies Based on Empirical Data.* Springer-Verlag, New York, 1982.

[532] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 1991.

[533] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies to their probabilities. *Theory Prob. Appl.*, pages 264–280, 1971.

[534] V. N. Vapnik and A. Y. Chervonenkis. Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory Prob. Appl.*, pages 532–553, 1981.

[535] S. Verdu. The capacity region of the symbol-asynchronous Gaussian multiple-access channel. *IEEE Trans. Inf. Theory*, pages 733–751, July 1989.

[536] S. Verdu. *Recent Progress in Multiuser Detection* (Advances in Communication and Signal Processing), Springer-Verlag, Berlin, 1989. [Reprinted in N. Abramson (Ed.), *Multiple Access Communications*, IEEE Press, New York, 1993.]

[537] S. Verdu. The exponential distribution in information theory. *Probl. Inf. Transm. (USSR)*, pages 86–95, Jan.–Mar. 1996.

[538] S. Verdu. Fifty years of Shannon theory. *IEEE Trans. Inf. Theory*, pages 2057–2078, Oct. 1998.

[539] S. Verdu. *Multiuser Detection.* Cambridge University Press, New York, 1998.

[540] S. Verdu and T. S. Han. A general formula for channel capacity. *IEEE Trans. Inf. Theory*, pages 1147–1157, July 1994.

[541] S. Verdu and T. S. Han. The role of the asymptotic equipartition property in noiseless source coding. *IEEE Trans. Inf. Theory*, pages 847–857, May 1997.

[542] S. Verdu and S. W. McLaughlin (Eds.). *Information Theory: 50 Years of Discovery*. Wiley–IEEE Press, New York, 1999.

[543] S. Verdu and V. K. W. Wei. Explicit construction of optimal constant-weight codes for identification via channels. *IEEE Trans. Inf. Theory*, pages 30–36, Jan. 1993.

[544] A. C. G. Verdugo Lazo and P. N. Rathie. On the entropy of continuous probability distributions. *IEEE Trans. Inf. Theory*, IT-24:120–122, 1978.

[545] M. Vidyasagar. *A Theory of Learning and Generalization*. Springer-Verlag, New York, 1997.

[546] K. Visweswariah, S. R. Kulkarni, and S. Verdu. Source codes as random number generators. *IEEE Trans. Inf. Theory*, pages 462–471, Mar. 1998.

[547] A. J. Viterbi and J. K. Omura. *Principles of Digital Communication and Coding*. McGraw-Hill, New York, 1979.

[548] J. S. Vitter. Dynamic Huffman coding. *ACM Trans. Math. Software*, pages 158–167, June 1989.

[549] V. V. V'yugin. On the defect of randomness of a finite object with respect to measures with given complexity bounds. *Theory Prob. Appl.*, 32(3):508–512, 1987.

[550] A. Wald. *Sequential Analysis*. Wiley, New York, 1947.

[551] A. Wald. Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, pages 595–601, 1949.

[552] M. J. Weinberger, N. Merhav, and M. Feder. Optimal sequential probability assignment for individual sequences. *IEEE Trans. Inf. Theory*, pages 384–396, Mar. 1994.

[553] N. Weiner. *Cybernetics*. MIT Press, Cambridge, MA, and Wiley, New York, 1948.

[554] T. A. Welch. A technique for high-performance data compression. *Computer*, 17(1):8–19, Jan. 1984.

[555] N. Wiener. *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. MIT Press, Cambridge, MA, and Wiley, New York, 1949.

[556] H. J. Wilcox and D. L. Myers. *An Introduction to Lebesgue Integration and Fourier Series*. R.E. Krieger, Huntington, NY, 1978.

[557] F. M. J. Willems. The feedback capacity of a class of discrete memoryless multiple access channels. *IEEE Trans. Inf. Theory*, IT-28:93–95, 1982.

[558] F. M. J. Willems and A. P. Hekstra. Dependence balance bounds for single-output two-way channels. *IEEE Trans. Inf. Theory*, IT-35:44–53, 1989.

[559] F. M. J. Willems. Universal data compression and repetition times. *IEEE Trans. Inf. Theory*, pages 54–58, Jan. 1989.

[560] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens. The context-tree weighting method: basic properties. *IEEE Trans. Inf. Theory*, pages 653–664, May 1995.

[561] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens. Context weighting for general finite-context sources. *IEEE Trans. Inf. Theory*, pages 1514–1520, Sept. 1996.

[562] H. S. Witsenhausen. The zero-error side information problem and chromatic numbers. *IEEE Trans. Inf. Theory*, pages 592–593, Sept. 1976.

[563] H. S. Witsenhausen. Some aspects of convexity useful in information theory. *IEEE Trans. Inf. Theory*, pages 265–271, May 1980.

[564] I. H. Witten, R. M. Neal, and J. G. Cleary. Arithmetic coding for data compression. *Commun. ACM*, 30(6):520–540, June 1987.

[565] J. Wolfowitz. The coding of messages subject to chance errors. *Ill. J. Math.*, 1:591–606, 1957.

[566] J. Wolfowitz. *Coding Theorems of Information Theory*. Springer-Verlag, Berlin, and Prentice-Hall, Englewood Cliffs, NJ, 1978.

[567] P. M. Woodward. *Probability and Information Theory with Applications to Radar*. McGraw-Hill, New York, 1953.

[568] J. Wozencraft and B. Reiffen. *Sequential Decoding*. MIT Press, Cambridge, MA, 1961.

[569] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering*. Wiley, New York, 1965.

[570] A. Wyner. A theorem on the entropy of certain binary sequences and applications II. *IEEE Trans. Inf. Theory*, IT-19:772–777, 1973.

[571] A. Wyner. The common information of two dependent random variables. *IEEE Trans. Inf. Theory*, IT-21:163–179, 1975.

[572] A. Wyner. On source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, IT-21:294–300, 1975.

[573] A. Wyner and J. Ziv. A theorem on the entropy of certain binary sequences and applications I. *IEEE Trans. Inf. Theory*, IT-19:769–771, 1973.

[574] A. Wyner and J. Ziv. The rate distortion function for source coding with side information at the receiver. *IEEE Trans. Inf. Theory*, IT-22:1–11, 1976.

[575] A. Wyner and J. Ziv. On entropy and data compression. *IEEE Trans. Inf. Theory*, 1991.

[576] A. D. Wyner. Capacity of the the band-limited Gaussian channel. *Bell Syst. Tech. J.*, 45:359–395, Mar. 1966.

[577] A. D. Wyner. Communication of analog data from a Gaussian source over a noisy channel. *Bell Syst. Tech. J.*, pages 801–812, May–June 1968.

[578] A. D. Wyner. Recent results in the Shannon theory. *IEEE Trans. Inf. Theory*, pages 2–10, Jan. 1974.

[579] A. D. Wyner. The wiretap channel. *Bell Syst. Tech. J.*, pages 1355–1387, 1975.

[580] A. D. Wyner. The rate-distortion function for source coding with side information at the decoder. II: General sources. *Inf. Control*, pages 60–80, 1978.

[581] A. D. Wyner. Shannon-theoretic approach to a Gaussian cellular multiple-access channel. *IEEE Trans. Inf. Theory*, pages 1713–1727, Nov. 1994.

[582] A. D. Wyner and A. J. Wyner. Improved redundancy of a version of the Lempel–Ziv algorithm. *IEEE Trans. Inf. Theory*, pages 723–731, May 1995.

[583] A. D. Wyner and J. Ziv. Bounds on the rate-distortion function for stationary sources with memory. *IEEE Trans. Inf. Theory*, pages 508–513, Sept. 1971.

[584] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Inf. Theory*, pages 1–10, Jan. 1976.

[585] A. D. Wyner and J. Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inf. Theory*, pages 1250–1258, Nov. 1989.

[586] A. D. Wyner and J. Ziv. Classification with finite memory. *IEEE Trans. Inf. Theory*, pages 337–347, Mar. 1996.

[587] A. D. Wyner, J. Ziv, and A. J. Wyner. On the role of pattern matching in information theory. *IEEE Trans. Inf. Theory*, pages 2045–2056, Oct. 1998.

[588] A. J. Wyner. The redundancy and distribution of the phrase lengths of the fixed-database Lempel–Ziv algorithm. *IEEE Trans. Inf. Theory*, pages 1452–1464, Sept. 1997.

[589] A. D. Wyner. The capacity of the band-limited Gaussian channel. *Bell Syst. Tech. J.*, 45:359–371, 1965.

[590] A. D. Wyner and N. J. A. Sloane (Eds.) *Claude E. Shannon: Collected Papers*. Wiley–IEEE Press, New York, 1993.

[591] A. D. Wyner and J. Ziv. The sliding window Lempel–Ziv algorithm is asymptotically optimal. *Proc. IEEE*, 82(6):872–877, 1994.

[592] E.-H. Yang and J. C. Kieffer. On the performance of data compression algorithms based upon string matching. *IEEE Trans. Inf. Theory*, pages 47–65, Jan. 1998.

[593] R. Yeung. *A First Course in Information Theory*. Kluwer Academic, Boston, 2002.

[594] H. P. Yockey. *Information Theory and Molecular Biology*. Cambridge University Press, New York, 1992.

[595] Z. Zhang, T. Berger, and J. P. M. Schalkwijk. New outer bounds to capacity regions of two-way channels. *IEEE Trans. Inf. Theory*, pages 383–386, May 1986.

[596] Z. Zhang, T. Berger, and J. P. M. Schalkwijk. New outer bounds to capacity regions of two-way channels. *IEEE Trans. Inf. Theory*, IT-32:383–386, 1986.

[597] J. Ziv. Coding of sources with unknown statistics. II: Distortion relative to a fidelity criterion. *IEEE Trans. Inf. Theory*, IT-18:389–394, 1972.

[598] J. Ziv. Coding of sources with unknown statistics. II: Distortion relative to a fidelity criterion. *IEEE Trans. Inf. Theory*, pages 389–394, May 1972.

[599] J. Ziv. Coding theorems for individual sequences. *IEEE Trans. Inf. Theory*, pages 405–412, July 1978.

[600] J. Ziv. Distortion-rate theory for individual sequences. *IEEE Trans. Inf. Theory*, pages 137–143, March 1980.

[601] J. Ziv. Universal decoding for finite-state channels. *IEEE Trans. Inf. Theory*, pages 453–460, July 1985.

[602] J. Ziv. Variable-to-fixed length codes are better than fixed-to-variable length codes for Markov sources. *IEEE Trans. Inf. Theory*, pages 861–863, July 1990.

[603] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Trans. Inf. Theory*, IT-23:337–343, 1977.

[604] J. Ziv and A. Lempel. Compression of individual sequences by variable rate coding. *IEEE Trans. Inf. Theory*, IT-24:530–536, 1978.

[605] J. Ziv and N. Merhav. A measure of relative entropy between individual sequences with application to universal classification. *IEEE Trans. Inf. Theory*, pages 1270–1279, July 1993.

[606] W. H. Zurek. Algorithmic randomness and physical entropy. *Phys. Rev. A*, 40:4731–4751, Oct. 15 1989.

[607] W. H. Zurek. Thermodynamic cost of computation, algorithmic complexity and the information metric. *Nature*, 341(6238):119–124, Sept. 1989.

[608] W. H. Zurek (Ed.) *Complexity, Entropy and the Physics of Information* (Proceedings of the 1988 Workshop on the Complexity, Entropy and the Physics of Information). Addison-Wesley, Reading, MA, 1990.

LIST OF SYMBOLS

INDEX

Elements of Information Theory, Second Edition, By Thomas M. Cover and Joy A. Thomas