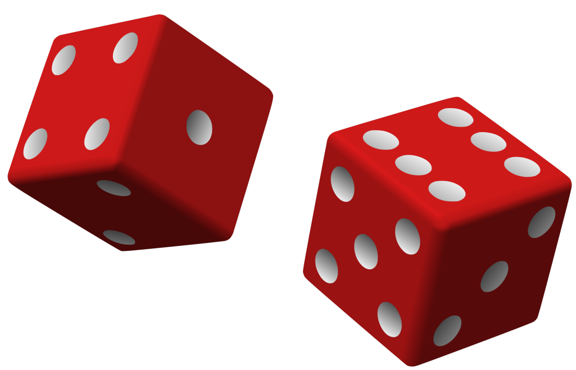
Information without rolling dice

Taehyung J. Lim
Massimo Franceschetti

Simons Conference, Austin, Texas, May 21, 2015



Shannon’s information theory

$$C = W \log \sqrt{1 + \mathsf{SNR}}$$

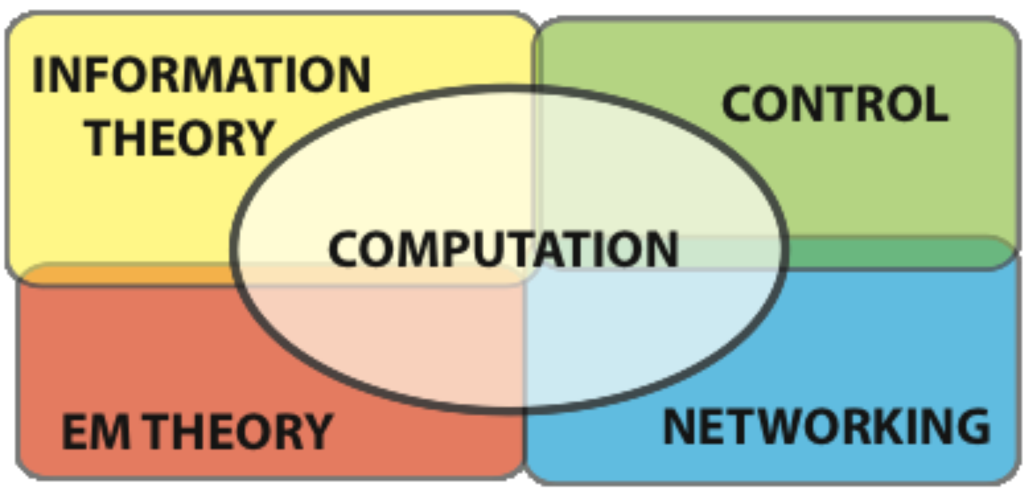
$$H = -\sum p_i \log p_i$$

Analogous notions in a deterministic setting?

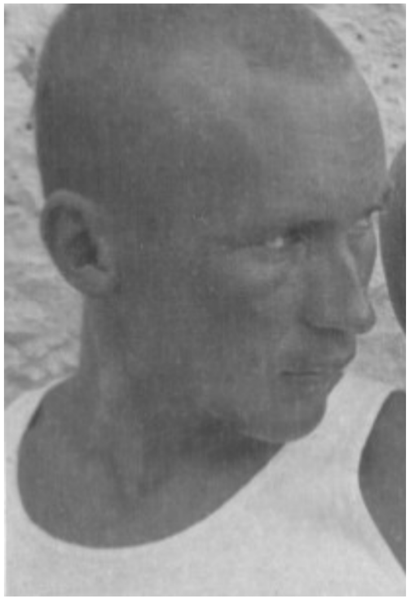
Why deterministic?

◇ Networked Control Systems (CPS)
◇ Nair (2013), Matveev and Savkin (2007)

◇Electromagnetic Wave Theory
◇Franceschetti (2015), Franceschetti Migliore and Minero (2009)



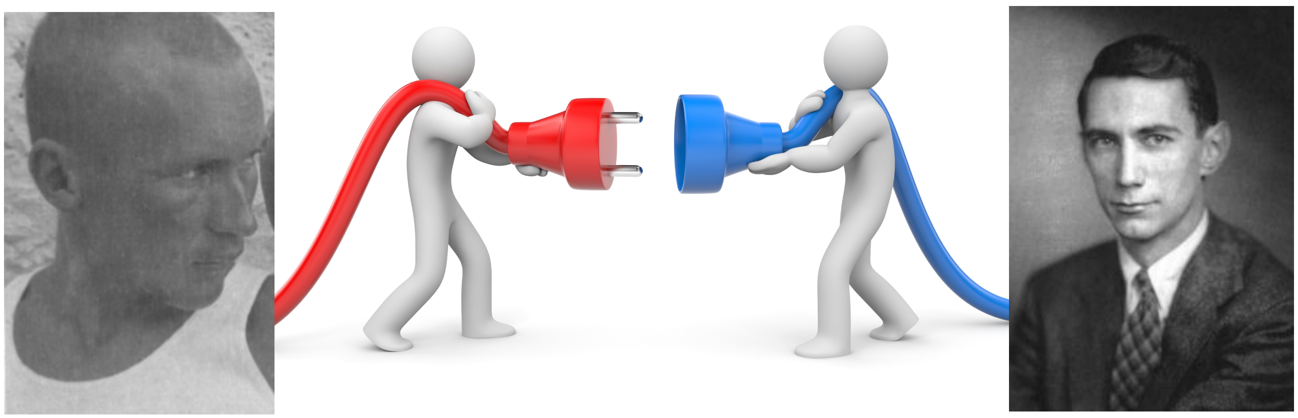
Deterministic setting

◇"*His mathematical intuition is remarkably precise*" A.N. Kolmogorov on Shannon's work

◇Influenced by Shannon's work he introduced (1956-1959) the concepts of entropy and capacity in the deterministic framework of functional approximation.

Deterministic vs Stochastic



◇He drew some connections in Kolmogorov (ISIT 1956), and Kolmogorov and Tikhomirov (AMS translation 1961, Appendix II), but limited the discussion "*at the level of analogy and parallelism*"

◇ At the time, the theory spectral concentration of bandlimited functions was not yet developed by Landau, Pollack, and Slepian (60s,70s)

◇Donoho (2000) describes connection between entropy and rate

Capacity

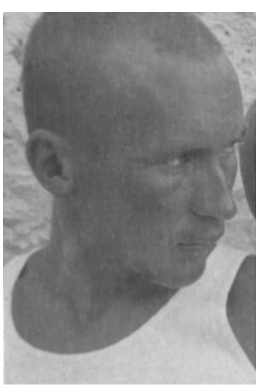
$$\int_{-\infty}^{\infty} f^2(t)dt \leq E$$

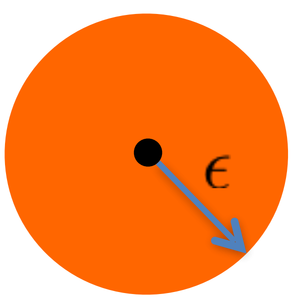
$f \in \mathscr{B}_\Omega$

$N_0 = \Omega T/\pi = 2WT$

Perturbation ϵ

$\text{SNR}_K = E/\epsilon^2$





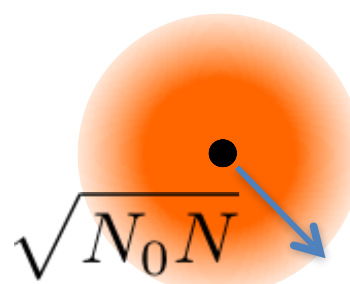
$$\int_{-T/2}^{T/2} f^2(t)dt \leq PN_0$$

$f \in \mathscr{B}_\Omega$

$N_0 = \Omega T/\pi = 2WT$

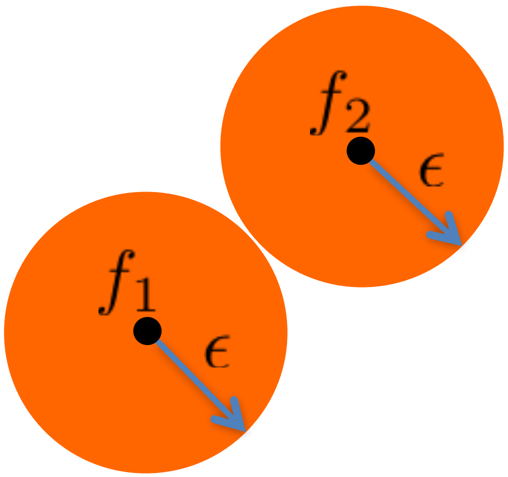
Gaussian std $= \sqrt{N}$

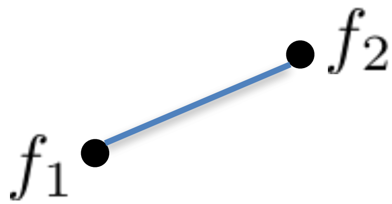
$\text{SNR}_S = P/N$

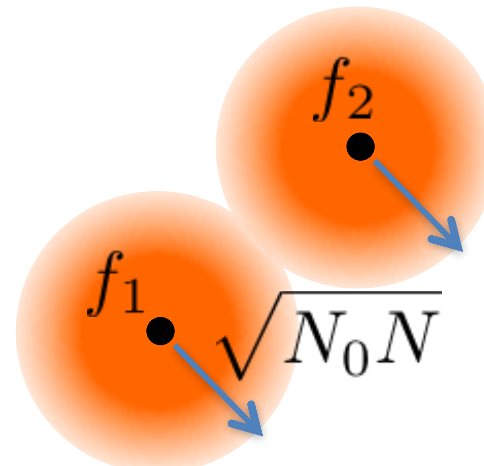


Capacity

$$\|f_1 - f_2\|^2 = \int_{-\frac{T}{2}}^{\frac{T}{2}} (f_1(t) - f_2(t))^2 \, dt$$

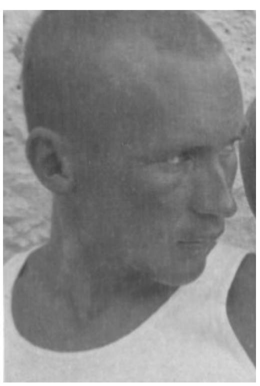




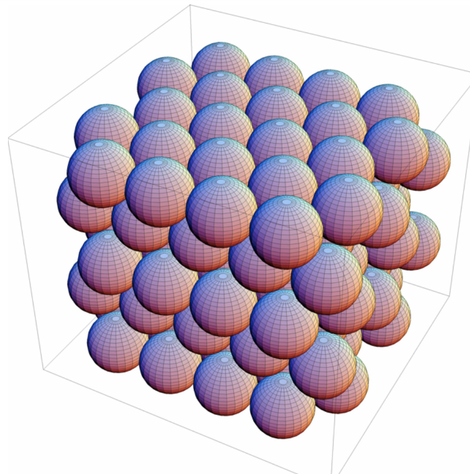


distinguishable

distinguishable w.h.p.



Capacity



$\mathcal{M}_{2\epsilon}(E) \equiv$ max nr. distinguishable points

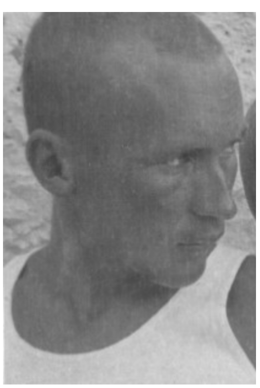
$C_{2\epsilon} = \log \mathcal{M}_{2\epsilon}(E)$ bits

$$\bar{C}_{2\epsilon} = \lim_{T\to\infty} \frac{\log \mathcal{M}_{2\epsilon}(E)}{T} \text{ bps}$$

$\mathcal{M}_N(P) \equiv$ max nr. points : $p_{\text{err}} < p_e$

$C(p_e) = \log \mathcal{M}_N(P)$ bits

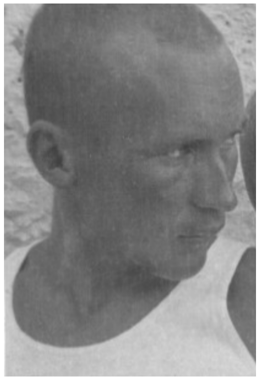
$$C = \lim_{T\to\infty} \frac{\log \mathcal{M}_N(P)}{T} \text{ bps}$$

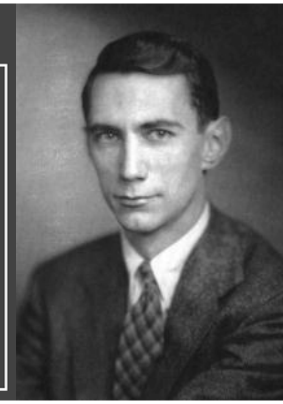


Comparison

◇The two models are driven by the same geometric intuition and have the same operational significance in terms of reliable communication with noise-perturbed signals

◇Shannon and Kolmogorov are playing the same game of packing billiard balls in high-dimensional space (and keeping the score in bits)





◇ Jagerman's bounds (1969-70)

$$0 \le \bar{C}_{2\epsilon} \le \frac{\Omega}{\pi} \log\left(1 + 2\sqrt{\mathsf{SNR}_K}\right) \ \text{bps}$$

$$C = \frac{\Omega}{\pi} \log(\sqrt{1 + \mathsf{SNR}_S}) \ \text{bps}$$

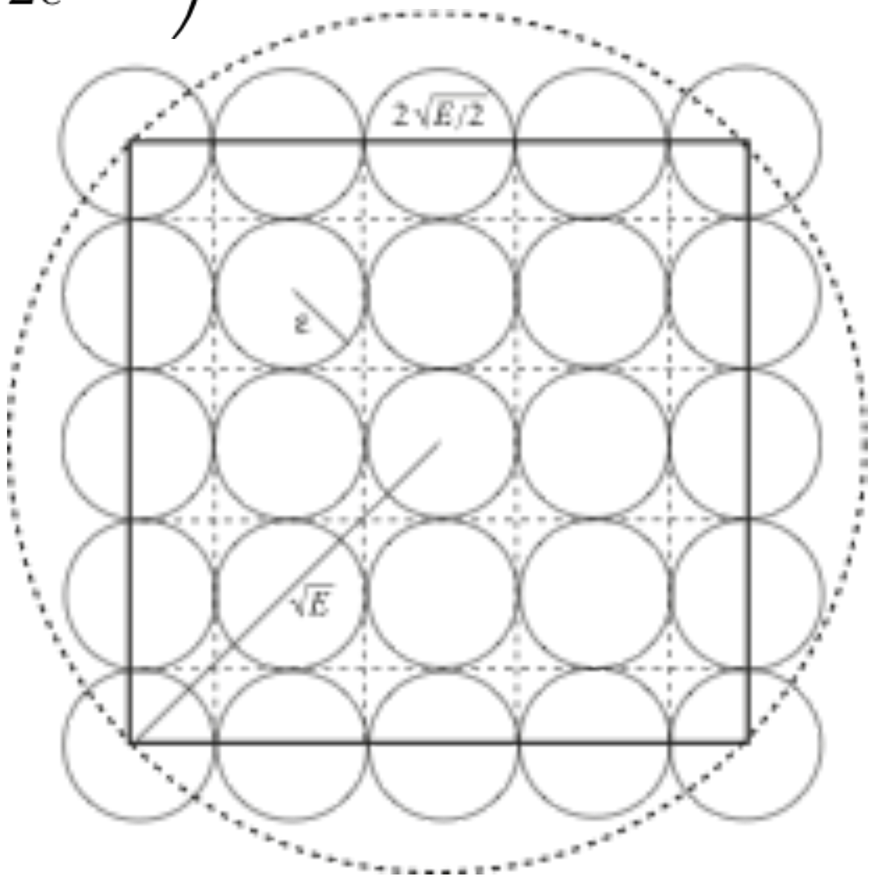
Lower bound

◇Both volumes of square and the ball tend to zero
◇Volume of square vanishes at a faster rate
◇Volume escapes on the boundary of the sphere

$$\# \text{ Balls} \geq \left(1 + \frac{2\sqrt{E/N_0}}{2\epsilon}\right)^{N_0} = (1 + \sqrt{\mathsf{SNR}_K/N_0})^{N_0}$$

$$C_{2\epsilon} \geq \sqrt{\mathsf{SNR}_K N_0}$$

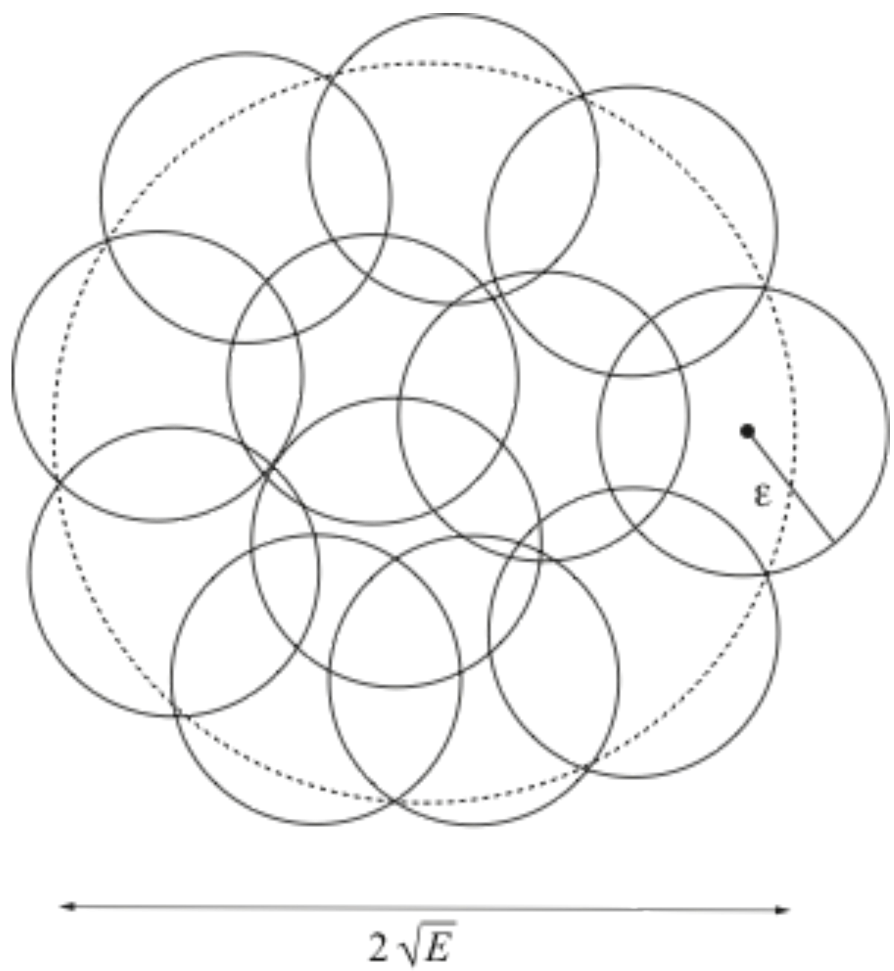
$$\bar{C}_{2\epsilon} \geq 0$$



Upper bound

◇Follows from Kolmogorov inequality $C_{2\epsilon} \leq H_\epsilon$

◇And upper bound on H_ϵ



◇ Lim and F. (Preprint)

$$\boxed{\frac{\Omega}{\pi}(\log\sqrt{\mathsf{SNR}_K}-1) \le \bar{C}_{2\epsilon} \le \frac{\Omega}{\pi}\log(1+\sqrt{\mathsf{SNR}_K/2})}$$

$$\mathsf{SNR}_K \gg 1: \frac{\Omega}{\pi}(\log\sqrt{\mathsf{SNR}_K}-1) \le \bar{C}_{2\epsilon} \le \frac{\Omega}{\pi}(\log\sqrt{\mathsf{SNR}_K}-1/2)$$

◇Tight up to half a bandwidth

◇Capacity scales linearly with bandwidth and logarithmically with SNR

◇Do not provide an explicit codebook, but rely on geometric properties of covering and packing

End of the game?

◇ *"What is the notion of an error exponent in a deterministic setting?"* Francois Baccelli at Allerton 2014.

Revisit the model

Allow for a certain amount of overlap in the packing of the balls

Signals in a codebook are (ϵ, δ) -distinguishable if the portion of space Δ where the received signal may lie and result in a decoding error is at most δ

$$\mathcal{M}_\epsilon(E) \equiv \text{max nr. points} : \Delta < \delta$$

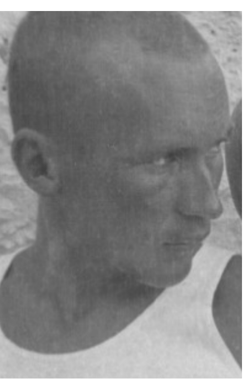
$$C_\epsilon^\delta = \log \mathcal{M}_\epsilon(E) \text{ bits}$$

$$\bar{C}_\epsilon^\delta = \lim_{T\to\infty} \frac{\log \mathcal{M}_\epsilon(E)}{T} \text{ bps}$$

$$\mathcal{M}_N(P) \equiv \text{max nr. points} : p_{\text{err}} < p_e$$

$$C(p_e) = \log \mathcal{M}_N(P) \text{ bits}$$

$$C = \lim_{T\to\infty} \frac{\log \mathcal{M}_N(P)}{T} \text{ bps}$$



◇ Lim and F. (Preprint)

$$\boxed{\frac{\Omega}{\pi}\log\sqrt{\mathrm{SNR}_K} \le \bar{C}^{\delta}_{\epsilon} \le \frac{\Omega}{\pi}\log(1+\sqrt{\mathrm{SNR}_K})}$$

◇Tight in high SNR

◇As in Shannon’s case, the capacity per unit time does not depend on the size of the error region

Error exponent

Let number of messages in the codebook

$$M = 2^{TR}$$

Size of error region is at most

$$\delta \leq 2^{-T\left(\frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K}-R\right)}$$

Error exponent in the deterministic model is

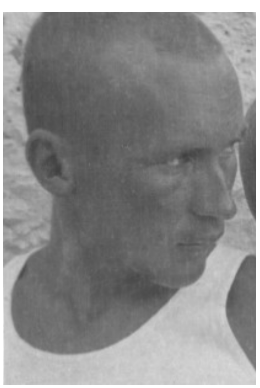
$$\mathsf{Er}(R) = \frac{\Omega}{\pi}\log\sqrt{\mathsf{SNR}_K} - R > 0$$

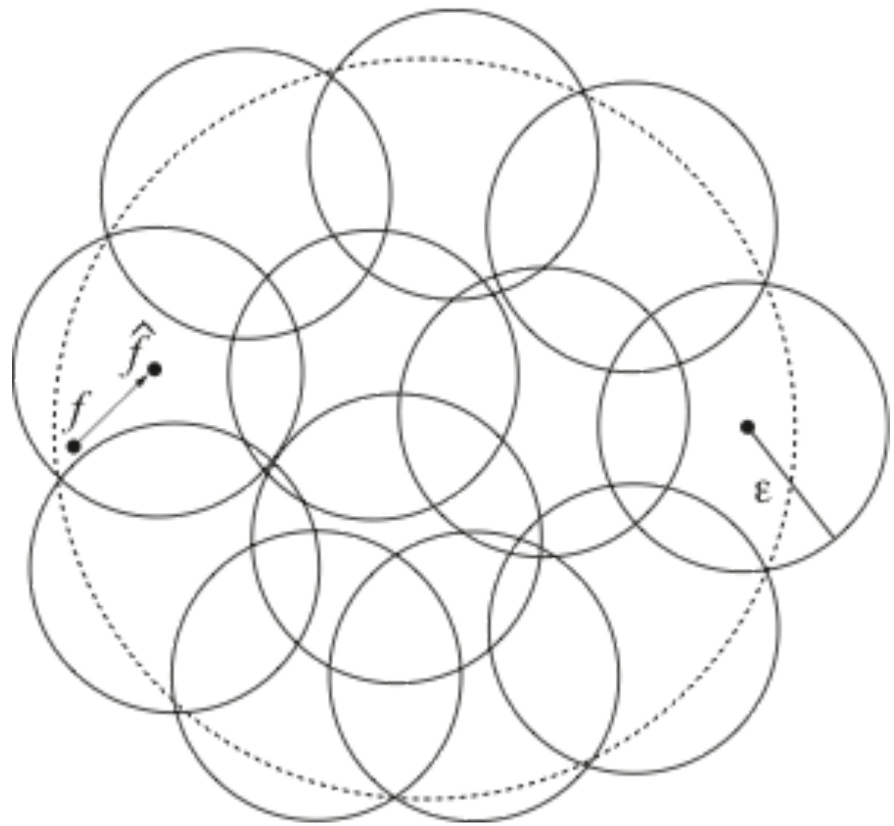
Entropy

◇Connections between deterministic and stochastic notions of entropy: Donoho (2000), Neuhoff and Gray (1998)

Kolmogorov entropy corresponds to the minimum number of bits that can represent any signal in the space with a given error constraint

$$\int_{-T/2}^{T/2}[f(t)-\hat{f}(t)]^2dt \le \epsilon^2$$





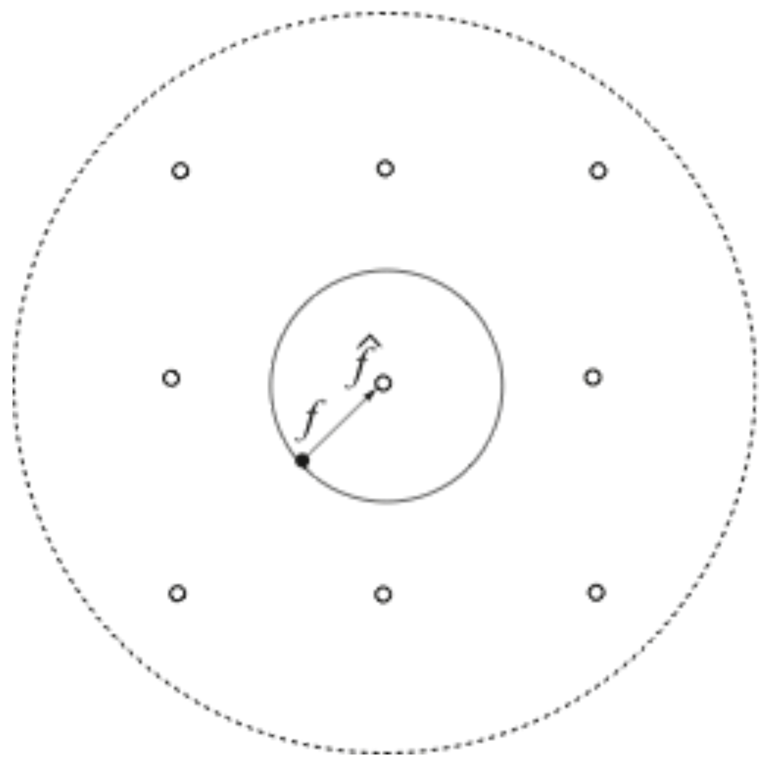
Entropy

Shannon's entropy corresponds to the average number of bits necessary to represent any quantized stochastic process of a given class, with uniform quantization on each coordinate of the space

If a stochastic process is represented by a codebook point with mean square error

$$\int_{-T/2}^{T/2} \mathbb{E}[\mathsf{f}(t) - \hat{\mathsf{f}}(t)]^2 dt$$

then the min number of bps that represent the process with a given accuracy is given by Shannon's rate distortion function



Entropy

$$\int_{-\infty}^{\infty} f^2(t)dt \leq E$$

$$\int_{-T/2}^{T/2} \mathbb{E}[\mathsf{f}(t)]^2 dt = PN_0$$

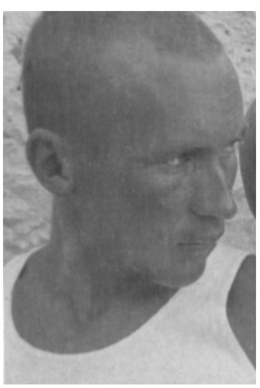
$$\int_{-T/2}^{T/2} [f(t) - \hat{f}(t)]^2 dt \leq \epsilon^2$$

$$\int_{-T/2}^{T/2} \mathbb{E}[\mathsf{f}(t) - \hat{\mathsf{f}}(t)]^2 dt \leq \sigma^2 N_0$$

$$H_\epsilon = \log L_\epsilon(E) \text{ bits}$$

$$\bar{H}_\epsilon = \lim_{T\to\infty} \frac{\log L_\epsilon(E)}{T} \text{ bps}$$

$$R_\sigma = \lim_{T\to\infty} \frac{\log L_\sigma(P)}{T} \text{ bps}$$



◇ Lim and F. (Preprint)

$$\boxed{\bar{H}_\epsilon = \frac{\Omega}{\pi} \log \sqrt{\mathsf{SNR}_K}} \qquad R_\sigma = \frac{\Omega}{\pi} \log \sqrt{\mathsf{SNR}_S}$$

◇ Jagerman (1969-70)

$$0 \leq \bar{H}_\epsilon \leq \frac{\Omega}{\pi} \log \left(1 + 2\sqrt{\mathsf{SNR}_K}\right)$$

It’s a draw!