Entropy

$$H(X) = \sum p(x) \log \frac{1}{p(x)} = E_{p(x)}\left[\log p(x)\right]$$

$$H(X,Y) = \sum_x \sum_y p(x,y) \log \frac{1}{p(x,y)}$$

Uses $p(x|y) = \frac{p(x,y)}{p(y)}$

$$H(X|Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x|y)}$$

$$H(X,Y) = H(X) + H(X|Y)$$

Source Coding

Non-singular $\rightarrow$ 1-1

Uniquely Decodable $\rightarrow$ All extension non-singular $\rightarrow$ Prefix / suffix / Neither

Prefix $\rightarrow$ No codeword is prefix of another

$\min_{l_1, l_2 \ldots l_n, \text{prefix}}$ $E[L(X)] = \sum p(x) l(x)$

Prefix Codes satisfy Kraft $\sum_i \frac{1}{D^{l_i}} \leq 1$

Optimal $\Rightarrow$
set each $l_i^* = \log_D \frac{1}{p_i}$

Shannon $\Rightarrow$ set each $l_i^* = \left\lceil \log \frac{1}{p_i} \right\rceil$

Huffman Code $\Rightarrow$ Asymptotically Optimal

$$d_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|$$

Property

i) $\Pr(X \neq Y) \geq d_{TV}(P, Q)$

$\exists$ coupling which achieves equality

ii) $d_{TV}((P_1, P_2, P_3 \ldots P_n), (Q_1, Q_2 \ldots Q_n)) \leq \sum_{i=1}^{n} d_{TV}(P_i, Q_i)$

K-L divergence

$$D_{KL}(P/Q) = \sum P(x) \log\left(\frac{Q(x)}{P(x)}\right) = E_P\left[\log \frac{Q(x)}{P(x)}\right]$$

i)

$$D_{KL}(P^n \| Q^n) = n\, D_{KL}(P/Q)$$

ii) Chain Rule

$$D_{KL}(P_{XY}/Q_{XY}) = D_{KL}(P_X/Q_X) + \underbrace{D_{KL}(P_{Y|X}, Q_{Y|X} | P_X)}_{\text{Conditional KL-div}}$$

iii) $D_{KL}(P_{Y|X} \| Q_{Y|X} | P_X) \quad D_{KL}(P_X P_{Y|X} \| Q_X Q_{Y|X})$

iv) $Y = f(X)$

$$\rightarrow D_{KL}(P_{Y|X} \| Q_{Y|X} | P_X) = \sum P_X \sum P_{Y|X=x} \ln \frac{P_{Y|X=x}}{Q_{Y|X=x}}$$

$$= 0$$

$$\rightarrow D_{KL}(P_Y \| Q_Y) \leq D_{KL}(P_X \| Q_X)$$

Add Comm KL Div

Pinsker →

$$d_{TV}^2(P,Q) \frac{2}{\log_2 e} \le D(P\|Q)$$

Hypothesis Testing

Type I error $\alpha = P(A^c) = 1 - P(A)$

Type 2 $\beta = Q(A)$

LRT

$$A_\tau = \left\{ x : \frac{P(x)}{Q(x)} \geq \tau \right\}$$

Neyman – Pearson

If $Q(A) = Q(A_\tau)$

$\Rightarrow$ $1 - P(A) \geq 1 - P(A_\tau)$

$P(A) \le P(A_\tau)$

$\rightarrow \alpha + \beta \geq 1 - d_{TV}(P,Q)$

Chernoff Bound

$$\Pr(|X - np| > \delta np) \le 2e^{-\frac{\delta^2 np}{3}} \qquad 0 < \delta < 1$$

Stein's lemma

For large n, $\exists$ a test such that

$$\alpha_n \le \varepsilon, \quad \beta_n \le e^{-n D(P||Q)} \quad \forall \varepsilon \in (0,1)$$

For any test, β_n cannot fall faster than ⬭, i.e.

$$\beta_n \geq C_1 e^{-n D(P||Q)}$$

Converse: $\alpha_n + \beta_n \geq 1 - n\, d_{TV}(P,Q) \geq 1 - \sqrt{\frac{n}{2} D_e(P||Q)}$

Error exponents

$$E_0 < e^{-E_0 n} \qquad E_1 < e^{-n E_1}$$

$$(0, D(P||Q)) \ \& \ (D(Q||P), 0)$$

Sanov's Thm: $E \to$ Convex set of distributions

$$\Pr(\hat{P} \in E) = e^{-n \min_{Q \in E} D_e(Q||P) + f(n)}$$

$$\Pr(\hat{P} \in E) \le (n+1)^{|X|} e^{-n \min_{Q \in E} D_e(Q||P)}$$

Channel Coding Theorem

Channel Capacity $C = \max_{p_x} I(X;Y)$

$I(X;Y) = H(X) - H(X|Y)$

→ $I(X;X) = H(X)$

→ ≥ 0

→ $I(X;Y|Z) = I(X|Z;Y) = H(X|Z) - H(X|YZ)$

$= H(Y|Z) - H(Y|XZ)$

→ $I(X;Y|Z) = 0$ iff $p(x,z|y) = p(x|y)\,p(z|y)$

X, Z cond. indep on Y

$X \to Y \to Z$ → Markov

→ $I(X_1, X_2, \ldots, X_n|Y) = \sum_{i=1}^{n} I(X_i;Y|X_1 \ldots X_{i-1})$

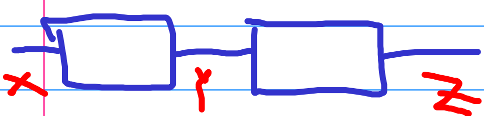
→ $H(X)$, $I(X;Y)$ given $P_{Y|X}$ are strictly concave in X

⇒ Unique maximum

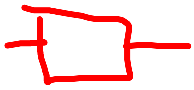
Rows are permutations of each other

$C(P_{Y|X}) = \max_{P(X)} H(Y) - H(\text{row})$

All col sums equal ⇒ X uniform ⇒ Y uniform ⇒ $\log_2|Y| - H_2(\text{row})$



$C_3 = \min(C_1, C_2)$



Product of Two channels!

$$C \leq C_1 + C_2,$$

Typical Set → Robust typicality

$$T_\varepsilon^{(n)}(x) = \{ x^n : |\pi(x, x^n) - p(x)| \le \varepsilon p(x) \}$$

For any $x^n \in T_\varepsilon^{(n)}(x)$

$$(1-\varepsilon) E[g(x)] \le \frac{1}{n} \sum_{i=1}^{n} g(x_i) \le (1+\varepsilon) E[g(x)]$$

$$\to 2^{-nH(x)(1+\varepsilon)} \le p(x^n = x^n) \le 2^{nH(x)(1+\varepsilon)}$$

$$\to |T_\varepsilon^{(n)}(x)| \le 2^{nH(x)(1+\varepsilon)}$$

$\to Pr(x^n \in T_\varepsilon^{(n)} x) \ge 1-\varepsilon$ for n large enough

$$\Rightarrow |T_\varepsilon^{(n)}(x)| \ge 2^{nH(x)}$$

Fano's inequality

$X, \hat{X} \sim \mathcal{X} = \{1, 2, \ldots, M\}$

$P_e = Pr(\hat{X} \neq X)$

$$H(X|\hat{X}) \le H_2(P_e) + P_e \log(M-1)$$

AWGN capacity

Restrict Power of X => $E[X^2] \le P$

More general

Constraint $E[g(x)] \le B$

$$C(P_{y|x}, B) = \max_{p(x) \sim E[g(x)] \le B} I(X;Y)$$

AWGN: $\frac{1}{2} \log\left(1 + \frac{P}{\sigma^2}\right)$

Parameter Estimation

$L(\theta, \hat{\theta}) \longrightarrow$ 0/1

1-norm

2-norm

$$R^n_{\hat{\theta}}(\theta) = E_{X^n \sim P(\theta)}\left[L(\theta, \hat{\theta})\right]$$

I Method of Mom:

Assume Sample mean = dist. mean

II MLE:

$$\hat{\theta}_{ML} = \arg\max_{\theta} L_\theta(x^n)$$

$$\text{Risk} = E\left[(\hat{\theta} - \theta)^2\right] = \text{Var}(\hat{\theta}) + (\text{Bias}(\hat{\theta}))^2$$

Unbiased est for variance => $\frac{\sum_i (x_i - \bar{\mu})^2}{n-1}$ ML -> Divide by n

$$\bar{\sigma}_{ML} = \frac{\sum_i (x_i - \bar{\mu})^2}{n} \qquad \bar{\mu}_{ML} = \frac{\sum_i x_i}{n} \qquad \bar{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$E[\bar{\sigma}] = E\left[(x_i - \bar{\mu})^2\right]$$

$$= E\left[(x_i - \mu)^2 + (\bar{\mu} - \mu)^2 - 2(x_i - \mu)(\bar{\mu} - \mu)\right]$$

$$= \left[\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right] = \boxed{\sigma^2\left(1 - \frac{1}{n}\right)}$$

$$\left[(x_i - \mu)(x_i - \mu) + (n-1)(x_i - \mu)(x_j - \mu)\right]$$

$$\text{Bias} = \boxed{-\frac{\sigma^2}{n}} \qquad \text{Unbiased} : \frac{\sum_i (x_i - \mu)^2}{n-1}$$

$x_i - \bar{\mu}$

Add Column MLE

Bayes Risk

$$R_{\hat{\theta}}(f(\theta)) = \int R_{\bar{\theta}}(\theta) f_\theta(\theta)\, d\theta$$

Bayesian: $\underset{\bar{\theta}}{\arg\min}\ R_{\bar{\theta}}^{(n)}(f(\theta))$

Bayesian Risk $\leq$ Minimax Risk

Minimax: $\underset{\bar{\theta}}{\arg\min}\ \underset{\theta}{\max}\ R_{\bar{\theta}}^{(n)}(\theta)$

Bayesian

$$\underbrace{P(X^n = x^n \mid \Theta = \theta)}_{\text{likelihood}} \; \underbrace{p_\Theta(\theta)}_{\text{prior}} = \underbrace{\Pr(\Theta = \theta \mid X^n = x^n)}_{\text{posterior}} \; \underbrace{P(X^n = x^n)}_{\text{indep of } \theta}$$

Posterior $\propto$ prior $\times$ likelihood

Estimate $\hat\theta = E[\theta \mid x^n] \rightarrow$ mean of posterior

Add Priors & Conj Priors

Posterior Mean optimal $\rightarrow$

Consider $\hat\theta \rightarrow$ estimator $\rightarrow$ some $g(x^n)$

$$\text{Avg-Loss} = \int (\hat\theta(x^n) - \theta)^2 \, p(X^n = x^n \mid \Theta = \theta) \, p(\Theta = \theta) \, dx^n \, d\theta$$

$$= \int \underbrace{(\hat\theta(x^n) - \theta)^2 \, p(\Theta = \theta \mid X^n = x^n) \, d\theta}_{\text{Minimized when } \hat\theta(x^n) = \text{Mean(Posterior)}}$$

Minimax

$$\text{Risk} \le \max_\theta E[(\hat\theta - \theta)^2] \text{ for any } \hat\theta$$

$$\text{Risk} \geq \min_g \underset{\theta \sim p_\theta(\theta)}{E}[(\hat\theta - \theta)^2] = E_\theta[(\hat\theta_{Bayes} - \theta)^2]$$

Cond 1

$$\int p_\theta(\theta)\, E\left[(\hat\theta_{Bayes}-\theta)^2\right] d\theta \geq E\left[(\hat\theta_{Bayes}-\theta)^2\right] \quad \forall \theta$$

$\hat\theta_{Bayes} \to$ Minmax

Pf:

$$\max_\theta E\left[(\hat\theta_{Bayes}-\theta)^2\right] \leq \text{Risk} \leq \max_\theta E\left[(\hat\theta_{Bayes}-\theta)^2\right]$$

$$\text{Risk} = \max_\theta E\left[(\hat\theta_{Bayes}-\theta)^2\right]$$

Cond 2

Risk indep of θ =)

Minmax for $x_i \sim Ber(p)$

$$\frac{X_1+\cdots+X_n+\sqrt{n}/2}{n+\sqrt{n}}$$

Le Cam

$$D_{KL}(P_Y, Q_Y) \leq D_{KL}(P_X, Q_X)$$