Infotheory for Statistics and Learning

Lecture 2

- Random transformations [PW:2.4]
- Distortion–rate and rate–distortion [PW:24,26],[CT10]
- Bounds [PW:26],[CT:10]
- Iterative computation [PW:5.6],[CT:10.8]

Random Transformation

Consider two measurable spaces $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{F})$, then a stochastic kernel from $\mathcal{X}$ to $\mathcal{Y}$ is a mapping $K(\cdot|\cdot)$ such that

1) For any fixed $x \in \mathcal{X}$, $K(\cdot|x)$ is a probability measure on $(\mathcal{Y}, \mathcal{F})$

2) For any fixed $F \in \mathcal{F}$, $K(F|\cdot) : \mathcal{X} \to \mathbb{R}$ is measurable

- For random variables $X : \Omega \to \mathcal{X} = \mathbb{R}$ and $Y : \Omega \to \mathcal{Y} = \mathbb{R}$, K defines a conditional distribution $P_{Y|X=x}$
- Also known as: random transformation, transition probability kernel, Markov kernel, channel

Given P_X on $(\mathbb{R}, \mathcal{B})$ and a kernel $P_{Y|X=x}(\cdot) = K(\cdot|x)$ we get

$$P_{XY}(E) = \int \left\{ \int \mathbb{1}\{(x,y) \in E\} dP_{Y|X=x} \right\} dP_X$$

on $(\mathbb{R}^2, \mathcal{B}^2)$, for $E \in \mathcal{B}^2$, and

$$P_Y(B) = \int \left\{ \int_B dP_{Y|X=x} \right\} dP_X$$

on $(\mathbb{R}, \mathcal{B})$ for $B \in \mathcal{B}$

Given P_X and $P_{Y|X=x}$ we say $P_{Y|X=x}$ induces P_Y and P_{XY}, notation:

$$X \overset{P_{Y|X}}{\to} Y \quad \text{or} \quad P_Y = P_{Y|X} \circ P_X$$

We also use $P_{XY} = P_{Y|X} \times P_X$

Distortion vs Information Rate

Consider describing a RV X with another variable Y through the transformation

$$X \overset{P_{Y|X}}{\to} Y$$

with resulting average distortion $E[d(X,Y)]$, for a given $d : \mathbb{R}^2 \to [0, \infty]$, and subject to an information constraint

$$I(X;Y) \leq R$$

To get the optimal kernel, solve

$$D(R) = \inf_{P_{Y|X}: I(X;Y) \leq R} E[d(X,Y)]$$

- The distortion–rate function of X (P_X)
- $D(R)$ is convex and non-increasing
- $D(R)$ is continuous on (R_0, ∞), $R_0 = \inf\{R : D(R) < \infty\}$

$D(R)$ has an inverse $R(D) = D^{-1}(R)$, which solves

$$R(D) = \inf_{P_{Y|X}: E[d(X;Y)] \le D} I(X,Y)$$

- The rate–distortion function of X (P_X)
- $R(D)$ is convex and non-increasing
- $R(D)$ is continuous on (D_0, ∞), $D_0 = \inf\{D : R(D) < \infty\}$

Generalizes to $X^n = (X_1, \dots, X_n)$ and $Y^n = (Y_1, \dots, Y_n)$:

$$D_n(R) = \inf_{P_{Y^n|X^n}: I(X^n;Y^n) \le R} \sum_{i=1}^{n} E[d(X_i, Y_i)]$$

$$R_n(D) = \inf_{P_{Y^n|X^n}: \sum_i E[d(X_i;Y_i)] \le D} I(X^n; Y^n)$$

And when the limits exist,

$$D_\infty(R) = \lim_{n\to\infty} \frac{1}{n} D_n(R), \quad R_\infty(D) = \lim_{n\to\infty} \frac{1}{n} R_n(D)$$

Gaussian X

For $\{X_i\}$ zero-mean stationary Gaussian with $\phi(k) = E[X_i X_{i-k}]$ and

$$\Phi(\omega) = \sum_k \phi(k) e^{-jk\omega}$$

and with $d(x, y) = (x - y)^2$, we get $(D, R_\infty(D)) = (d_\theta, r_\theta)$ where

$$d_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\theta, \Phi(\omega)\} d\omega$$
$$r_\theta = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max\left\{0, \log \frac{\Phi(\omega)}{\theta}\right\} d\omega$$

for $0 \leq \theta \leq \operatorname{ess\,sup} \Phi(\omega)$

For iid we get $\Phi(\omega) = E[X_i^2] = \sigma^2$, $\operatorname{ess\,sup} \Phi(\omega) = \sigma^2$ and

$$d_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\theta, \sigma^2\} d\omega = \theta$$
$$r_\theta = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max\left\{0, \log \frac{\sigma^2}{\theta}\right\} d\omega = \frac{1}{2} \log \frac{\sigma^2}{\theta}$$

That is,

$$R_\infty(D) = R(D) = \frac{1}{2} \log \frac{\sigma^2}{D}, \quad 0 \leq D \leq \sigma^2$$

For $E[X_i X_{i-k}] = \sigma^2 \rho^k$, $0 < \rho < 1$, we instead get

$$R_\infty(D) = \frac{1}{2} \log \frac{\sigma^2(1 - \rho^2)}{D}, \quad D \leq \frac{1 - \rho}{1 + \rho}$$

and otherwise the parametric expression

Testing for Optimality

Is $P^*_{Y|X}$ optimal?

Let $P^*_Y = P^*_{Y|X} \circ P_X$ and find $P^*_{X|Y}$ via $P_X = P^*_{X|Y} \circ P^*_Y$

If $E[d(X,Y^*)] \le D$ and for any other P_{XY} with $E[d(X,Y)] \le D$

$$E_{P_{XY}}\left[\log \frac{dP_{X|Y^*}}{dP_X}\right] \ge I(X;Y^*)$$

then $R(D) = I(X;Y^*)$

Conversely, suppose $I(X;Y^*) = R(D)$, then if for any $P_{X|Y^*}$ and for any P_{XY} that satisfies $E[d(X,Y)] \le D$, we have $P_Y \ll P^*_Y$ and $I(X;Y) < \infty$, then the inequality above holds

Maximum $h(X)$

X abs. continuous with pdf $f(x)$ and $\int x^2 f(x)dx = \sigma^2$

Let

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}$$

Then

$$0 \le \int f(x) \ln \frac{f(x)}{g(x)} dx = \frac{1}{2} \ln 2\pi\sigma^2 + \frac{1}{2\sigma^2} \int x^2 f(x)dx - h(X)$$

Thus

$$h(X) \le \frac{1}{2} \ln 2\pi e\sigma^2 \quad \text{[nats]}$$

with $=$ iff $f(x) = g(x)$

Entropy Bound on MMSE

For abs. continuous (X, Y) with pdf $f(x, y)$, let

$$\Delta(y) = E[(X - \hat{x}(Y))^2 | Y = y], \quad \hat{x}(y) = E[X | Y = y], \quad \Delta = E[\Delta(Y)]$$

and set

$$g(x|y) = \frac{1}{\sqrt{2\pi\Delta(y)}} \exp\left(-\frac{1}{2\Delta(y)}(x - \hat{x}(y))^2\right)$$

We have

$$\begin{aligned} 0 &\leq \int f(x|y) \ln \frac{f(x|y)}{g(x|y)} dx \\ &= \int f(x|y) \ln f(x|y) dx - \int f(x|y) \ln g(x|y) dx \end{aligned}$$

And thus,

$$\begin{aligned} h(X|Y = y) &\leq \frac{1}{2} \ln 2\pi\Delta(y) + \frac{1}{2\Delta(y)} \int (x - \hat{x}(y))^2 f(x|y) dx \\ &= \frac{1}{2} \ln 2\pi e \Delta(y) \end{aligned}$$

Consequently,

$$h(X|Y) \leq \frac{1}{2} E[\ln 2\pi e \Delta(Y)] \leq \frac{1}{2} \ln 2\pi e E[\Delta(Y)] = \frac{1}{2} \ln 2\pi e \Delta$$

with $=$ iff $f(x|y) = g(x|y)$

That is

$$E[(X - \hat{x}(Y))^2] \geq \frac{1}{2\pi e} e^{2h(X|Y)}$$

(with $h(X|Y)$ in nats) with $=$ iff (X, Y) are jointly Gaussian

For $n > 1$ dimensions:

For P_X on $(\mathbb{R}^n, \mathcal{B}^n)$, $X = (X_1, \dots, X_n)$, with $E[X^T X] = R$

$$h(X) \le \frac{1}{2}\log(2\pi e)^n |R| = \frac{n}{2}\log 2\pi e + \frac{1}{2}\mathrm{Tr}\log R$$

($|R|$ = determinant, Tr = trace) with $=$ only for X Gaussian

And for P_{XY} on $(\mathbb{R}^n \times \mathbb{R}^m, \mathcal{B}^n \times \mathcal{B}^m)$

$$|E[(X - \hat{x}(Y))^T (X - \hat{x}(Y))]| \ge \frac{1}{(2\pi e)^n} 2^{2h(X|Y)}$$

with $=$ iff (X, Y) are jointly Gaussian and $n = m$

Bounds on $R(D)$

For X abs. continuous with pdf $f(x)$ and $\int x^2 f(x)dx = \sigma^2$

Define $P_{Y|X}$ by $Y = \alpha^2 X + \alpha W$ where $W \sim \mathcal{N}(0, D)$ and independent of X, and

$$\alpha = \sqrt{\frac{\sigma^2 - D}{\sigma^2}}, \quad D \le \sigma^2$$

Then $E[Y^2] = \sigma^2 - D$, $E[(X - Y)^2] = D$ and

$$I(X;Y) = h(Y) - h(Y|X) \le \frac{1}{2}\log 2\pi e(\sigma^2 - D) - \frac{1}{2}\log 2\pi e \alpha^2 D$$
$$= \frac{1}{2}\log\frac{\sigma^2}{D}$$

Thus with $d(x, y) = (x - y)^2$,

$$R(D) \le \frac{1}{2}\log\frac{\sigma^2}{D}$$

for any X with $E[X^2] = \sigma^2$ and $=$ only for Gaussian

Consider $P_{Y|X}$, with pdf $f(y|x)$, such that $E[(X-Y)^2] \le D$

For $P_Y = P_{Y|X} \circ P_X$ we have $P_X = P_{X|Y} \circ P_Y$ where $P_{X|Y}$ has pdf $f(x|y)$ and P_Y has pdf $f(y)$. Set

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}, \quad h(x|y) = \frac{1}{\sqrt{2\pi D}} e^{-(x-y)^2/(2D)}$$

Then

$$I(X;Y) + D(f(x)\|g(x))$$
$$= \int \left\{ \int f(x|y) \ln \frac{f(x|y)}{h(x|y)} dx \right\} f(y) dy + \iint f(x,y) \ln \frac{h(x|y)}{g(x)} dx dy$$

$$\ge \iint f(x,y) \ln \frac{\sqrt{\sigma^2}\exp(-(x-y)^2/(2D))}{\sqrt{D}\exp(-x^2/(2\sigma^2))} dx dy \ge \frac{1}{2} \ln \frac{\sigma^2}{D} \quad \text{[nats]}$$

Thus for X abs. continuous with pdf $f(x)$ and $\int x^2 f(x) dx = \sigma^2$

$$\frac{1}{2} \log \frac{\sigma^2}{D} - D(f(x)\|g(x)) \le R(D) \le \frac{1}{2} \log \frac{\sigma^2}{D}$$

with $=$ only for $f(x) = g(x)$

The lower bound is tight for small D, i.e.

$$\lim_{D \to 0} \frac{R(D)}{\frac{1}{2} \log \frac{\sigma^2}{D} - D(f(x)\|g(x))} = 1$$

Iterative Computation of $R(D)$

For $(\mathbb{R}, \mathcal{B}, P_X)$ we get $P_Y = P_{Y|X} \circ P_X$ for a given $P_{Y|X}$ and we can find $P_{Y|X}$ for a given P_Y

Note that the $R(D)$-problem is convex, thus we can minimize

$$\int \left\{ \int \log \frac{dP_{Y|X=x}}{dP_Y} dP_{Y|X=x} \right\} dP_X + \lambda E[d(X,Y)]$$

over $P_{Y|X}$, $\lambda > 0$, but complicated since P_Y depends on $P_{Y|X}$

Consider instead minimizing

$$\int \left\{ \int \log \frac{dP_{Y|X=x}}{dQ_Y} dP_{Y|X=x} \right\} dP_X + \lambda E[d(X,Y)]$$

for fixed Q_Y and then over Q_Y for fixed $P_{Y|X}$

Abs. continuous: $P_X \to f(x)$, $P_{Y|X} \to f(y|x)$ and $Q_Y \to q(y)$

For fixed $q(y)$ the optimal $f(y|x)$ is

$$f(y|x) = \frac{q(y)e^{-\lambda d(x,y)}}{\int q(y)e^{-\lambda d(x,y)} dy}$$

and for fixed $f(y|x)$ the optimal $q(y)$ is

$$q(y) = \int f(x) f(y|x) dx$$

Pick an initial $q(y)$ and solve for $f(y|x)$

Solve for a new $q(y)$; Solve for a new $f(y|x)$; Iterate

- Has a unique stationary point generating the optimal $f(y|x)$
- Obvious modification to discrete variables