Supplemental Notes

EE503 Week 10 Dr. Franzke

HW # 10

① Handout

② Leon-Garcia 4.90-4.94, 4.147 – 4.149

Topics

① Transforms of random variables

$$f_Y \overset{1-1}{=} f_X \cdot \left| \frac{dx}{dy} \right|$$

② Information Theory

$$H(Y|X) \leq H(Y)$$

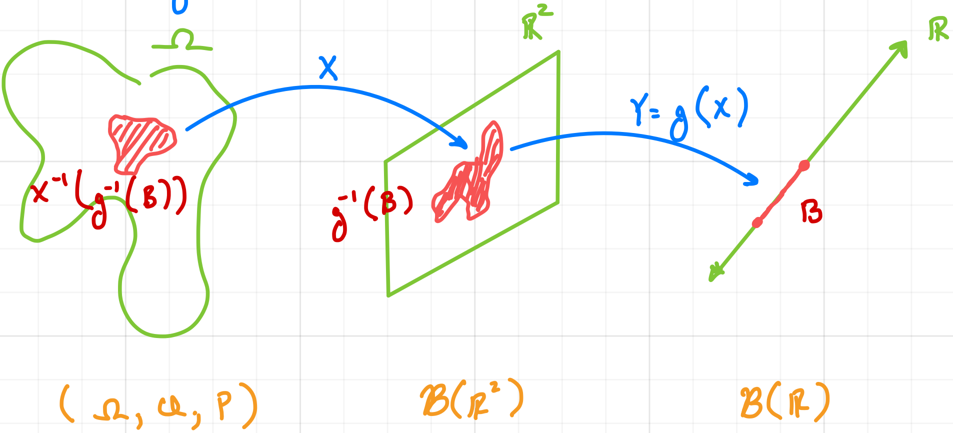
③ Uncertainty and Fuzzy sets

$$P[A] = E[a]$$

↑ fuzzy set.

Transformed Random Variables

$Y = g(X)$



★

Thrm: $f_Y(y) = \sum_{k=1}^{n} f_X(x_k) \cdot \left| \frac{dx}{dy} \right|_{x = x_k}$ ↙ roots of $Y = g(X)$

(from change of variable theorem)

Ex: Linear $Y = g(X) = aX + b$ w/ $a \neq 0$.

$\therefore g$ is 1-1.

$$\therefore f_Y(y) \overset{1-1}{=} f_X(x) \cdot \left| \frac{dx}{dy} \right|$$

$X = \frac{Y-b}{a}$ since $a \neq 0$ $\qquad \therefore \frac{dx}{dy} = \frac{1}{a}.$

$$\therefore \left| \frac{dx}{dy} \right| = \frac{1}{|a|}$$

$$\boxed{\therefore f_Y(y) = \frac{1}{|a|} \cdot f_X\left(\frac{Y-b}{a} \right)}$$

Vector case: $\underset{n\times 1}{Y} = \underset{n\times 1}{b} + \underset{n\times n}{A}\,\underset{n\times 1}{X}$ and A non-singular

$\therefore A^{-1}$ exists

$$\therefore f_Y(y) = f_X(x) \cdot \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|$$

where $X = (x_1(y_1, \dots, y_n), \dots, x_n(y_1, \dots, y_n))$

$$= \frac{f_X(A^{-1}(y-b))}{|\det(A)|}$$

Ex: Exponential $Y = e^X$ (1-1)

$\therefore X = \ln Y$ since $Y > 0$

$\therefore \left|\frac{dx}{dy}\right| = \left|\frac{1}{y}\right| = \frac{1}{y}$ since $y > 0$

$\therefore f_Y(y) = \frac{1}{y} \cdot f_X(\ln y)$

Say $X \sim N(\mu, \sigma_x^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_x}\right)^2}$

$\therefore$ log-normal: $f_Y(y) = \frac{1}{y\sqrt{2\pi}\,\sigma_x} e^{-\frac{1}{2}\left(\frac{\ln y - \mu_x}{\sigma_x}\right)^2}$. ($Y > 0$: 0 else)

$\therefore E_Y[Y] = e^{\mu_x + \sigma_x^2/2}$

$V_Y[Y] = e^{2\mu_x + \sigma_x^2}\left(e^{\sigma_x^2} - 1\right)$

$\therefore Y \sim \text{LogNormal}(\mu_x, \sigma_x^2) \Longrightarrow \ln Y \sim N(\mu_x, \sigma_x^2)$

and $\ln Y^k \sim N(k\mu_x, k^2\sigma_x^2)$

Further: If $Y_1 \sim \text{LogNormal}(\mu_1, \sigma_1^2)$ and

$Y_2 \sim \text{LogNormal}(\mu_2, \sigma_2^2)$ are independent.

then: $Y_1 \cdot Y_2 \sim \text{LogNormal}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Ex: Square $Y = aX^2$ for $a > 0$ ($\because$ not 1-1)

$y < 0 \longrightarrow$ no real roots

$\therefore f(y) = 0$ if $y < 0$

$y > 0 \longrightarrow$ 2 solutions : $x = \pm\sqrt{\frac{y}{a}}$

$$g(X) = Y = aX^2.$$

$$\therefore g'(x) = \frac{dy}{dx} = 2ax = \pm 2a\sqrt{\frac{y}{a}} = \pm 2\sqrt{ay}$$

$$\therefore |g'(x)| = 2\sqrt{ay}$$

$$\therefore f(y) = \frac{f_x(x_1)}{|g'(x_1)|} + \frac{f_x(x_2)}{|g'(x_2)|}$$ ($= 0$ if $y < 0$)

$$= \frac{1}{2\sqrt{ay}}\left(f_x(\sqrt{y/a}) + f_x(-\sqrt{y/a})\right)$$

2-D case follows from Change of Variable Theorem

$$\iint_A f(x,y)\,dx\,dy = \iint_{A^*} f(x(u,v), y(u,v)) \left| \frac{\partial(x,y)}{\partial(u,v)} \right| du \cdot dv$$

Sampling from a distribution

Ex: Want to draw iid $X_1, \dots, X_n \sim \exp(\theta)$

$$\left(X \sim \exp(\theta) : f(x) = \frac{1}{\theta} e^{-x/\theta} \right)$$

Trick:
- draw r.s. from $U[0,1]$.
- back-solve for $X \sim f(x)$

If X has closed form F^{-1} ★ (inverse CDF)

Else: Use Accept-Reject (or other algorithm, e.g. Box-Muller for $N(0,1)$)

Review

g^{-1} exists iff g is 1-1 and ONTO

- $g: X \to Y$ is ONTO: $g(X) = Y$.

$$\forall y \in Y \ \exists x \in X : y = g(x)$$

- g is 1-1 iff

$$g(x_1) = g(x_2) \Longrightarrow x_1 = x_2 \quad \forall x_1, x_2.$$

$$\left(x_1 \neq x_2 \longrightarrow g(x_1) \neq g(x_2) \right)$$

Ex: $g(x) = e^x$. Say $g(x_1) = g(x_2)$

$$\therefore e^{x_1} = e^{x_2} \qquad \therefore x_1 = \ln e^{x_1} = \ln e^{x_2} = x_2 \qquad \therefore \text{1-1.}$$

Continuous CDF w/ F^{-1})

★ **Thrm:** $Y = F(X) \sim U[0,1]$. (F^{-1} exists)

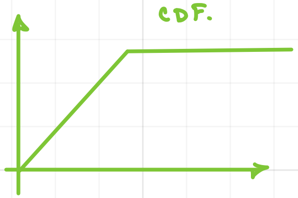
Prf: $F_Y(y) \overset{def}{=} P[Y \le y] = P[\{\omega \in \Omega : Y(\omega) \le y\}]$

$$\overset{hyp}{=} P[F_X(X) \le y]$$

$$\overset{1-1}{=} P[X \le F_X^{-1}(y)]$$

$$\overset{def\ CDF}{=} F_X(F_X^{-1}(y))$$

$$\overset{1-1}{=} y$$



$\therefore f_Y(y) = \frac{d}{dy} F_Y = \frac{dy}{dy} = 1.$

$\therefore Y \sim U(0,1)$ QED.

$\therefore$ turn it around.

continuous CDF

Thrm: $U \sim U[0,1]$ and $X \sim F$ then $X = F^{-1}(U)$ has CDF F.

Prf: $F_U(u) = u$ $\forall u \in (0,1)$ since $U \sim U[0,1]$.

$\therefore P(X \le x) = P(F^{-1}(U) \le x)$

$$= P(FF^{-1}(U) \le F(x))$$

$$= P(U \le F(x))$$

$$= F(x)$$

QED.

Ex: Find $n=3$ r.s. (iid) draws from $\exp(2)$

Want: $X_1, X_2, X_3 \sim f(x) = \frac{1}{2} e^{-x/2}$ $\quad (x>0)$

Sol: $Y = F(x)$ uniform from Theorem

∴ in general: $U = F(x) = 1 - e^{-x/\theta} \in (0,1)$ $\quad$ $(\exp(2))$

Backsolve: $\quad$ ∴ $X = -\theta \cdot \ln(1-U) \sim \exp(\theta)$

$u_k = 1 - e^{-x/2}$ $\quad (\exp(2))$ $\quad k=1,2,3.$

Pick "random number" in $u[0,1]$ for u_k:

pick $u_1 = 0.10480 \Longrightarrow$ $\quad$ ∴ $x_1 = 0.222$

realization

$u_2 = 0.24130 \Longrightarrow$ $\quad$ ∴ $x_2 = 0.552$

$u_3 = 0.22368 \Longrightarrow$ $\quad$ ∴ $x_3 = 0.506$

∴ iid draws: $\{0.222, 0.552, 0.506\} \sim \exp(2)$

☆ **Method:** Draw iid $X_1, \ldots, X_n$ from $X \sim f(x)$. If X continuous and F^{-1} exists

① $U = F(X) \sim u[0,1]$

② "Randomly" pick $u_k \in (0,1)$

③ Back-solve: $x_k = F^{-1}(u_k)$

Ex: Draw iid (standard Cauchy) $Z_1, \dots, Z_n \sim C(0,1)$

$$f(z) = \frac{1}{\pi} \cdot \frac{1}{1+z^2} \quad \longleftrightarrow \quad \text{continuous pdf.}$$

$$\therefore \; F(z) = \int_{-\infty}^{z} f(x)\,dx = \frac{1}{\pi} \tan^{-1} z + \frac{1}{2}.$$

$$\therefore \; U = F(Z) \sim U(0,1)$$

$$u_k = \frac{1}{\pi} \tan^{-1} z_k + \frac{1}{2}$$

$$\therefore \; z_k = \tan\left(\pi\left(u_k - \tfrac{1}{2}\right)\right) \sim C(0,1)$$

SαS: $\alpha \in (0,2)$

$\alpha = 1 \longrightarrow$ Cauchy

$\alpha = 2 \longrightarrow$ Normal

only closed forms.

C.F. $\phi(\omega) = e^{-\gamma|\omega|^{\alpha}}$

Q: what about $X_1, \dots, X_n \sim N(\mu, \sigma_x^2)$

Problem: no closed form F^{-1} $\therefore$ use Box-Müller

Generating Normal realizations (Box-Müller, 1958)

Thrm: Pick iid $Y_1 \sim U(0,1)$ and $Y_2 \sim U(0,1)$ then

X_1 and X_2 are iid $N(0,1)$ if

$$X_1 = \sqrt{-2 \cdot \ln Y_1}\, \cos(2\pi Y_2) \qquad X_2 = \sqrt{-2 \ln Y_1} \cdot \sin(2\pi Y_2)$$

Pf The above transform $(X_1, X_2) = g(Y_1, Y_2)$ is 1-to-1 and maps the open unit-square $(0,1)^2$ ONTO $\mathbb{R}^2$ except for sets involving $X_1 = 0$ or $X_2 = 0$ (have zero-probability)

∴ Can use $f(x_1, x_2) \overset{1-1}{=} |J| \cdot f(y_1, y_2)$ with solutions:

$$Y_1 = \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \qquad Y_2 = \frac{1}{2\pi} \tan^{-1}\left(\frac{x_2}{x_1}\right)$$

with $X_1 \neq 0$ and $X_2 \neq 0$

∴ The Jacobian is

$$J = \det \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix}$$

$$= \det \begin{vmatrix} (-x_1)\exp\left(-\frac{x_1^2+x_2^2}{2}\right) & (-x_2)\exp\left(-\frac{x_1^2+x_2^2}{2}\right) \\ \frac{-x_2/x_1^2}{2\pi\left(1+x_2^2/x_1^2\right)} & \frac{1/x_1}{2\pi\left(1+x_2^2/x_1^2\right)} \end{vmatrix}$$

$$= \frac{-\left(1+x_2^2/x_1^2\right)\exp\left(-\frac{x_1^2+x_2^2}{2}\right)}{2\pi\left(1+x_2^2/x_1^2\right)} = -\frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right)$$

But iid $Y_k \sim U(0,1) \Longrightarrow f(y_1, y_2) \overset{ind}{=} f(y_1) \cdot f(y_2) \overset{U(0,1)}{=} 1 \cdot 1 = 1$ on $(0,1)^2$

$$\therefore f(x_1, x_2) \overset{1-1}{=} |J| \, f(y_1, y_2) \overset{Y_k \sim U(0,1)}{=} |J|$$

$$= \frac{1}{2\pi}\exp\left(-\frac{x_1^2+x_2^2}{2}\right) = \left(\frac{1}{\sqrt{2\pi}} e^{-x_1^2/2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-x_2^2/2}\right)$$

$= f(x_1) \cdot f(x_2)$ ∴ iid $X_k \sim N(0,1)$ QED.

Monte Carlo Simulation

"Monte-Carlo" casino in Monaco

2-types

① iid M.C.

- generate (simulate) random samples (iid)
- Use LLN: $\overline{Y}_n \xrightarrow{mopd} (b-a)E_x[X]$

② "MCMC" Markov Chain MC (EE517)

- Correlated samples (from a Markov process)
  - ↳ converge to equilibrium pdf π_e
- Problem with "Burn-in" and time to convergence
- Often use Gibbs Sampler (software: R or BUGS)
  - ↳ Bayesian inference

Ex: Monte Carlo Integration

Q: how to evaluate $\int_a^b g(x)\,dx$ if g continuous but DON'T KNOW anti-derivative G?

A: IID MC estimate

① $$\int_a^b g(x)\,dx = \frac{b-a}{b-a}\int_a^b g(x)\,dx \qquad (a<b)$$

$$= (b-a)\int_a^b g(x)\cdot\left(\frac{1}{b-a}\right)dx$$

$$\overset{X\sim U[a,b]}{=} (b-a)\cdot E_X[g(X)].$$

② Compute: $Y_k \triangleq (b-a)\cdot g(X_k)$ ← simulate

from iid draws $X_k \sim U[a,b]$

$$\therefore \overline{Y_n} = \frac{1}{n}\sum_{k=1}^{n} Y_k = \frac{(b-a)}{n}\sum_{k=1}^{n} g(X_k)$$

$$\xrightarrow[P]{LLN} (b-a)\,E_X[g(X)] = \int_a^b g(x)\,dx$$

∴ consistent.

Gaussian Random Vectors

Defn: $\underset{n\times 1}{X} \sim N(\underset{n\times 1}{\mu_x}, \underset{n\times n}{K_{xx}})$ iff

$$f(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\mathrm{Det}(K_{xx})}} \cdot \exp\left(-\frac{1}{2}\underset{1\times n}{(X-\mu_x)^T}\underset{n\times n}{K_{xx}^{-1}}\underset{n\times 1}{(X-\mu_x)}\right)$$

if K_{xx}^{-1} exists where $K_{xx} = E_x[(X-\mu_x)(X-\mu_x)^T]$.

Defn: $\underset{n\times 1}{X}$ and $\underset{p\times 1}{Y}$ are Jointly Gaussian (JG) iff

$$Z = \left(\frac{X}{Y}\right) \sim N(\mu_z, K_{zz})$$

$$f(z) = \frac{1}{(2\pi)^{n+p/2}\sqrt{\mathrm{Det}(K_{zz})}} \cdot \exp\left(-\frac{1}{2}(z-\mu_z)^T K_{zz}^{-1}(z-\mu_z)\right)$$

$$\mu_z = \left(\frac{\mu_x}{\mu_y}\right) \qquad K_{zz} = \left(\begin{array}{c|c} K_{xx} & K_{xy} \\ \hline K_{yx} & K_{yy} \end{array}\right)$$

$\therefore$ X and Y independent if uncorrelated

Defn: If $\underset{n\times 1}{X}$ and $\underset{p\times 1}{Y}$ are JG: the Conditional Gaussian:

$$f(x|y) = \frac{1}{(2\pi)^{n/2}\sqrt{\det K_{x|y}}} \cdot \exp\left(-\frac{1}{2}(X-E[X|Y=y])^T K_{x|y}^{-1}(X-E[X|Y=y])\right)$$

$$E[X|Y] = \underset{n\times 1}{\mu_x} + \underset{n\times p}{K_{xy}}\,\underset{p\times p}{K_{yy}^{-1}}\,\underset{p\times 1}{(Y-\mu_y)}$$

scalar case $\left(= \mu_x + \rho_{xy}\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right)$

$$K_{x|y} = K_{xx} - K_{xy}K_{yy}^{-1}K_{yx}$$

scalar case $\left(= \sigma_x^2(1-\rho_{xy}^2)\right)$

Defn: Wishart Distribution W (k-dimensional χ^2).

$$\text{iid } \underset{n\times 1}{X_1}, \dots, \underset{n\times 1}{X_k} \sim N\left(\underset{n\times 1}{\mu_x}, \underset{n\times n}{K_{xx}}\right)$$

$$\therefore \underset{n\times n}{W} = \sum_{j=1}^{k} \underset{n\times 1}{(X_j - \mu_x)} \underset{1\times n}{(X_j - \mu_x)^T} \text{ is Wishart.}$$

$$f_W(w) = \frac{\exp\left(-\frac{1}{2}\mathrm{Trace}\left(K_{xx}^{-1} W\right)\right)\left(\mathrm{Det}(W)\right)^{\frac{1}{2}(k-n-1)}}{2^{kn/2}\left(\mathrm{Det}(K_{xx})\right)^{k/2} \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma\left(\frac{k+1-j}{2}\right)}$$

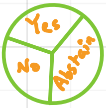
used in models of sample covariance.

3 Types of Uncertainty

① Entropic (Probabilistic Information Theory)

$$H(Y|X) \leq H(Y)$$

② Belief (Dempster-Shafer) Theory



③ Fuzzy (Vagueness)

$$F(A) = \mathcal{E}(A, A^c) \quad \text{and} \quad P(A) = E[a].$$

① Entropy.

$p = (p_1, \dots, p_n) \in \text{Simplex}$ → $p_k \geq 0$ and $\sum_{k=1}^{n} p_k = 1$

Unconditional Entropy

$$H(p) = -\sum_{k=1}^{n} p_k \cdot \log_2 p_k$$

$\therefore\ 0 \leq H(P) \leq \log_2 n.$

$$H(x) = -\sum_{x} p(x) \cdot \log_2 p(x)$$

Conditional Entropy

$$H(Y|X) = \sum_{x} p(x) \cdot H(Y|X=x)$$

$$= -\sum_{x} p(x) \sum_{y} p(y|x) \cdot \log_2 p(y|x)$$

$$= -\sum_{x}\sum_{y} p(x,y) \cdot \log_2 p(y|x)$$

$$= -E_{XY}\left[\log_2 p(Y|X)\right]$$

Mutual Information $I(X,Y)$

$$I(X,Y) = \text{"distance"}\left(\overset{\text{joint}}{p(x,y)}, \overset{\text{independent}}{p(x)p(y)}\right)$$

$$= \sum_{x}\sum_{y} p(x,y) \cdot \log_2 \frac{p(x,y)}{p(x)p(y)}$$

Kullback-Leibler "divergence"

$$= \sum_{x}\sum_{y} p(x,y) \cdot \log_2 \frac{p(x|y)}{p(x)}$$

$$= -\sum_{x}\left(\sum_{y} p(x,y)\right)\log_2 p(x) + \sum_{x}\sum_{y} p(x,y) \cdot \log_2 p(x|y)$$

$$= -\sum_x p(x)\cdot\log_2 p(x) - \left(-\sum_x\sum_y p(x,y)\log_2 p(x|y)\right)$$

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

Similarly: $I(X,Y) = H(Y) - H(Y|X)$

$p(x) > 0$
$p(y) > 0$

Thrm: $I(X,Y) \geq 0$ ($= 0 \longleftrightarrow$ X and Y independent)

Pf:

$$-I(X,Y) = -\sum_x\sum_y p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_x\sum_y p(x,y)\cdot\log_2\frac{p(x)p(y)}{p(x,y)}$$

$$= E_{XY}\left[\log_2\frac{P(X)P(Y)}{P(X,Y)}\right].$$

Jensen's Inq $\leq \log_2 E_{XY}\left[\frac{P(X)P(Y)}{P(X,Y)}\right]$ since logorithm concave

$$= \log_2\sum_x\sum_y \cancel{p(x,y)}\cdot\frac{p(x)p(y)}{\cancel{p(x,y)}}$$

$$= \log_2\left(\sum_x p(x)\right)\left(\sum_y p(y)\right)$$

$$= \log_2 1$$

$$= 0$$

$\therefore I(X,Y) \geq 0$ QED.

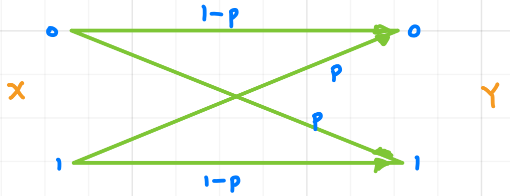
Thrm: $\boxed{H(Y|X) \leq H(Y)}$

"Conditioning REDUCES entropy."

Defn: Channel capacity C:

$$C = \max_{p(x)} I(X,Y)$$ in "bits" (binary digits)

Binary Symmetric Channel (BSC)



$p = P[\text{bit flip}]$

$$\therefore I(X,Y) = H(Y) - H(Y|X)$$

$$= H(Y) - \sum_x p(x) \cdot H(Y|X=x)$$

$$\overset{BSC}{=} H(Y) - \sum_x p(x) \cdot H(p)$$

↖ $H(p) = -p \cdot \log_2 p - (1-p)\log_2(1-p)$

$$= H(Y) - H(p)$$

$$\leq 1 - H(p)$$ since Y is a binary r.v.

$\therefore C = 1 - H(p)$ bits. ← since uniform pdf (input) maximizes $H(p)$

Gaussian Channel

- $Y_i = X_i + Z_i$

  independent

- $Z_i \sim N(0, N)$
- finite average signal power $\quad E_X[X^2] \le P.$

$$\therefore\ C = \max_{E[X^2]\le P} I(X,Y) = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right)$$

Further assume:

- Band-limited $X(t)$ with bandwidth W
- $Y(t) = \big(X(t) + Z(t)\big) * h(t)$

  convolution

  impulse response of ideal bandpass filter. (LTI)

$$\therefore\ \boxed{C = W\cdot\log_2\left(1+\frac{P}{N_0 W}\right)}$$ bits per second.

for white Gaussian (thermal) noise with power spectral density $\frac{N_0}{2}$

$Z \sim N\left(0, \frac{N_0}{2}\right)$

But $\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$

linear function of P.

$$\therefore\ \lim_{W\to\infty} C = \frac{P}{N_0}\cdot\log_2 e$$ bits per second

$\therefore$ infinite bandwidth $\not\Rightarrow$ infinite capacity

Thrm: (Asymptotic Equipartition, AEP) If iid $X_1,\dots,X_n \sim p(x)$:

$$-\frac{1}{n}\cdot \log_2 p(X_1,\dots,X_n) \xrightarrow{p} H(X)$$

$$\therefore\ p(\underbrace{X_1,\dots,X_n}_{\uparrow\ \text{random draw}}) \approx 2^{-nH(x)}$$

"Typical" sequence

Prf:

$$-\frac{1}{n}\log_2 p(x_1,\dots,x_n) = -\frac{1}{n}\sum_{k=1}^{n} \underbrace{\log_2 p(x_k)}_{\text{iid}}$$

$$\xrightarrow[p]{\text{WLLN}} -E_x[\log_2 p(X)].$$

$$\text{But } -E_x[\log_2 p(X)] = E_x\left[\log_2 \frac{1}{p(x)}\right]$$

$$= -\sum_x p(x)\cdot \log_2 p(x)$$

$$= H(X)$$

QED.

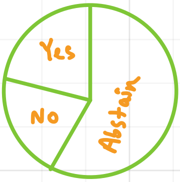
AEP underlies most proofs of channel capacity theorems
(i.e. reliable transmission rate R achievable if $R<C$)

② Belief Theory

Arthur Dempster (1967)
Glen Shafer (1970s)

"Dempster-Shafer" Theory

- allows one to ABSTAIN.
- upper and lower probabilities



Defn: Mass assignment $m : 2^X \longrightarrow [0,1]$ such that

① $m(\emptyset) = 0$

② $\sum_{A \in 2^X} m(A) = 1$ $X = \{x_1, \dots, x_n\}$.

∴ unit mass over 2^X (not sample space X)

∴ in general:
- $m(X) < 1$
- $m(A^c) \neq 1 - m(A)$
- $A \subset B \cancel{\Rightarrow} m(A) \leq m(B)$

∴ get back discrete probability if all sets are singleton sets

$$\mathrm{Bel}(A) = \mathrm{Belief}(A) = \sum_{B : B \subset A} m(B)$$

↑ all evidence B that supports A

$$\therefore \mathrm{Bel}(A) + \mathrm{Bel}(A^c) \leq 1$$

$$\mathrm{Pl}(A) = \mathrm{Plausibility}(A) = \sum_{B : A \cap B \neq \emptyset} m(B) = 1 - \mathrm{Bel}(A^c) \geq \mathrm{Bel}(A)$$

↑ all evidence B that fails to refute A.

$$\therefore \mathrm{Pl}(A) + \mathrm{Pl}(A^c) \geq 1.$$

Möbius Inversion Lemma gives:

$$m(A) = \sum_{B: B \subset A} (-1)^{c(A-B)} \cdot Bel(B)$$

where $c(A) = \sum_k a_k = \#\ 1s$

Ex: $X = \{X_1, X_2\} = \{H, T\}$

$\therefore 2^X = \{\emptyset, \{H\}, \{T\}, X\}.$

One of uncountably many mass assignments

$m(H) = 0.4$ $m(T) = 0\ 5$ $m(X) = 0.1$

Total uncertainty in D-S. theory

$\longleftrightarrow$ $m(X) = 1$ ("the keys are somewhere")

Avoids carving up X into probabilistic halves
(Prosecuter's Fallacy)

$m(X) = 1 \Longrightarrow m(A) = 0$ if $A \neq X$

Pick any $A \neq X$

$\therefore$ ① $Bel(A) = 0$ since $m(B) = 0$ for all $B \subset A$.

$\therefore$ Believe in no hypothesis (keys in room A)

② $Pl(A) = 1 + 0 + 0 + \cdots + 0 = 1$ since $X_n \cap A \neq \emptyset$ if $A \neq \emptyset$.

$\therefore$ All hypotheses are (maximally) plausible.

Thrm: (Dempster's Rule of Combination) m_1 and m_2 are independent (i.e. independent "experts")

$$\therefore\ m_{1,2}(A) = \frac{\sum\limits_{B}\sum\limits_{\substack{C \\ B\cap C = A}} m_1(B)\cdot m_2(C)}{1-D} \quad \text{for "discord"}$$

↑ combined belief

$$D = \sum_{B}\sum_{\substack{C \\ B\cap C=\emptyset}} m_1(B)\cdot m_2(C) \quad \leftarrow \text{measure of expert disagreement.}$$

Prf:

$$1 = 1\cdot 1 = \left(\sum_{B\in 2^X} m_1(B)\right)\left(\sum_{C\in 2^X} m_2(C)\right)$$

$$= \sum_{B\in 2^X}\sum_{C\in 2^X} m_1(B)\cdot m_2(C)$$

one way to partition

$$= \sum_{B}\sum_{\substack{C \\ B\cap C=\emptyset}} m_1(B)\, m_2(C) + \sum_{B}\sum_{\substack{C \\ B\cap C\neq\emptyset}} m_1(B)\cdot m_2(C)$$

$$= \underbrace{\sum_{B}\sum_{\substack{C \\ B\cap C=\emptyset}} m_1(B)\cdot m_2(C)}_{D} + \sum_{A\in 2^X}\left(\sum_{B}\sum_{\substack{C \\ B\cap C=A}} m_1(B)\cdot m_2(C)\right)$$

QED.

Note: problems arise if D is large

③ Fuzz (Vagueness)

Defn: $A \subset X$ is a Fuzzy Set iff

$I_A : X \longrightarrow [0,1]$ ← multivalued range

$\longleftrightarrow \quad a : X \longrightarrow [0,1]$

↑ a is "membership" or "FIT" value

↗ fuzzy unit.

$a(x) = \text{Degree}(x \in A)$

$A^c \longleftrightarrow a^c(x) = 1 - a(x)$

$A \cap B \longleftrightarrow \min(a(x), b(x))$

$A \cup B \longleftrightarrow \max(a(x), b(x))$

$\therefore$ A is properly fuzzy iff $A \cap A^c \neq \phi$

$A \cup A^c \neq X$

midpoint set $A \longleftrightarrow a(x) \triangleq \frac{1}{2}$

$\therefore A = A \cap A^c = A \cup A^c = A^c$

if finite $X = \{x_1, \dots, x_n\}$:

fuzzy power set $I^X = [0,1]^n$ ("fuzzy cube")

$\therefore$ Binary sets $A \in \{0,1\}^n \longleftrightarrow$ Vertex of $[0,1]^n$

↑ Boolean n-cube

Ex: $X = \{x_1, x_2\}$ $A = (a_1, a_2) = (\frac{1}{3}, \frac{3}{4})$

$$X = (\overset{x_1}{1}, \overset{x_2}{1})$$
$$\phi = (0, 0)$$
$$A = (\tfrac{1}{3}, \tfrac{3}{4})$$

"fit vector"

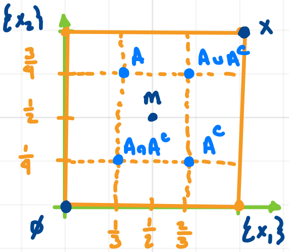
$$A^c = (\tfrac{2}{3}, \tfrac{1}{4}) \qquad A \cap A^c = (\tfrac{1}{3}, \tfrac{1}{4}) \qquad A \cup A^c = (\tfrac{2}{3}, \tfrac{3}{4})$$

$\neq \phi$ $\neq X$

$$M = (\tfrac{1}{2}, \tfrac{1}{2})$$

"Sets as points"

$A \in [0,1]^2$



Fuzzy Counting Measure $c(A)$:

$$c(A) = \begin{cases} \sum_x a(x) & \text{discrete case} \\ \int_X a(x)\,dx & \text{continuous case} \end{cases}$$

$A = (\frac{1}{3}, \frac{3}{4})$ $\therefore c(A) = \frac{1}{3} + \frac{3}{4} = \frac{13}{12} \notin \mathbb{Z}^+$

not an integer in general.

Fuzzy Equality:

$$\mathcal{E}(A, B) = \frac{c(A \cap B)}{c(A \cup B)}$$

Fuzziness (Vagueness) F:

$$F(A) = \frac{c(A \cap A^c)}{c(A \cup A^c)}$$

$$\therefore \; F(A) = \mathcal{E}(A, A^c)$$

Ex: $F(A) = \dfrac{\frac{1}{3} + \frac{1}{4}}{\frac{2}{3} + \frac{3}{4}} = \dfrac{7}{17}$

$$F(m) = 1$$

$F(B) = 0$ if B binary

Subsethood (Fuzzy Inclusion) S:

$$S(A, B) = \text{Degree}(A \subset B) = \frac{c(A \cap B)}{\underset{\uparrow}{c(A)}} \quad \text{if } c(A) > 0$$

resembles $P(B|A)$

$$\therefore \; F(A) = S(\underset{\text{"whole"}}{A \cup A^c}, \underset{\text{"part"}}{A \cap A^c})$$

∴ "whole in part"

Thm: $\boxed{\underset{\text{Fuzziness}}{F(A)} = \underset{\text{Subsethood}}{S(A \cup A^c, A \cap A^c)} = \underset{\text{Equality}}{\mathcal{E}(A, A^c)}}$

Ex: Suppose $\exists$ probability measure P such that $P \triangleq F$

(i.e. fuzz is "disguised" probability)

$\therefore \; \exists$ binary set $A \subset X$ (such as X itself by CAT)

$$0 < P(A) \overset{\text{hyp}}{=} F(A) = S(A \cup A^c, A \cap A^c) \overset{A \text{ binary}}{=} S(X, \emptyset) \overset{X \text{ is not a subset of } \emptyset \text{ if } X \neq \emptyset.}{=} 0$$

$\therefore \; 0 < 0 \quad \rightarrow \leftarrow$ contradiction

∴ There is no such P.

If $X = \{x_1, \ldots, x_n\}$ $\quad c(X) = n$

but $c(A) = \frac{n}{n} \cdot c(A) = n \cdot \frac{c(A \cap X)}{c(X)} = n \cdot S(X, A)$

∴ counting reduces to scaled subsethood

Relative frequency $\frac{n_A}{n}$:

$$\frac{n_A}{n} = \frac{\text{\# successes}}{\text{\# trials}} = \frac{c(A \cap X)}{c(X)} = S(X, A)$$

if A binary $\quad$ ∴ reduces to subsethood

Probability of a Fuzzy Event. (Zadeh, 1968)

Suppose $a : X \longrightarrow [0,1]$ is Borel-measureable

$$\therefore \boxed{P(A) = E[a]} \qquad (P(A) = E_X[I_A])$$

$$= \sum_x p(x)\, a(x) \qquad \left(= \int_x f(x)\, a(x)\, dx < \infty\right)$$

∴ In general: $P(A \cap A^c) > 0$

$P(A \cup A^c) < 1$

Ex: $P = \left(\overset{x_1}{\frac{1}{3}}, \overset{x_2}{\frac{2}{3}}\right)$

$A = \left(\frac{1}{3}, \frac{3}{4}\right)$

$A \cap A^c = \left(\frac{1}{3}, \frac{1}{4}\right)$

$A \cup A^c = \left(\frac{2}{3}, \frac{3}{4}\right)$

∴ $P(A) = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{3}{4} = \frac{11}{18}$

$P(A \cap A^c) = \frac{5}{18} > 0$

$P(A \cup A^c) = \frac{13}{18} < 1$