

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

$$f_y = f_x \cdot \left| \frac{dx}{dy} \right|$$

② Information Theory

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

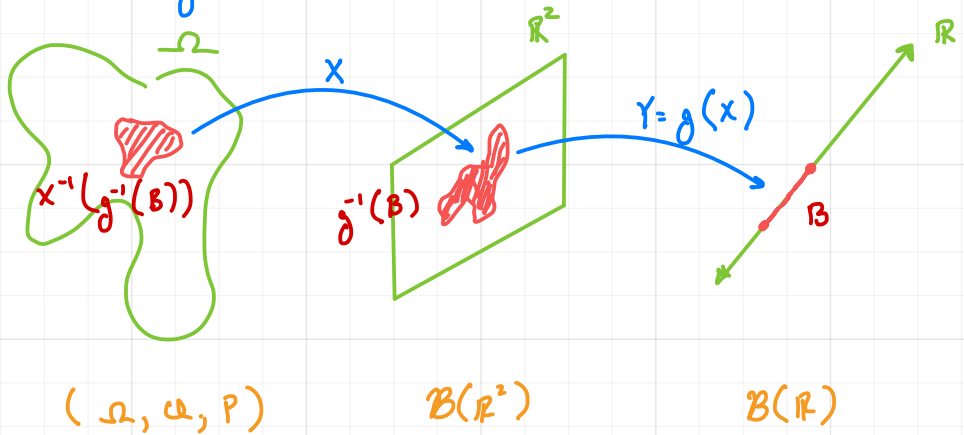
③ Uncertainty and Fuzzy sets

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

↑ fuzzy set.

Transformed Random Variables

$$Y = g(X)$$



Thm: $f_Y(y) = \sum_{k=1}^n f_X(x_k) \cdot \left| \frac{dx}{dy} \right|_{x=x_k}$ ↙ 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.

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

$$X = \frac{Y-b}{a} \text{ since } a \neq 0$$

$$\therefore \frac{dx}{dy} = \frac{1}{a}$$

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

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

Vector case: $Y = b + AX$ and A non-singular
 $\therefore A^{-1}$ exists

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

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

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

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

$$\therefore X = \ln Y \text{ since } Y > 0$$

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

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

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

$$\therefore \text{log-normal: } f_Y(y) = \frac{1}{y\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{\ln y - \mu_x}{\sigma_x}\right)^2}$$

($y > 0$ else)

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

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

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

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

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 $a > 0$ ($\therefore$ not 1-1)

$y < 0 \longrightarrow$ no real roots

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

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

$$g(x) = Y = ax^2.$$

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

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

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

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

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 $X_1, \dots, X_n \sim \exp(\theta)$

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

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

- $g: X \rightarrow Y$ is ONTO: $g(x) = y$.

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

- g is 1-1 iff

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

$$(x_1 \neq x_2 \implies g(x_1) \neq g(x_2))$$

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

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

★ Continuous CDF w/ F^{-1}

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

Prf: $F_Y(y) \stackrel{\text{def}}{=} P[Y \leq y] = P[\{\omega \in \Omega : Y(\omega) \leq y\}]$

$$\stackrel{\text{hyp}}{=} P[F_X(X) \leq y]$$

$$\stackrel{\text{I-1}}{=} P[X \leq F_X^{-1}(y)]$$

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

$$\stackrel{\text{I-1}}{=} y$$

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

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

QED.

$\therefore$ turn it around.

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

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

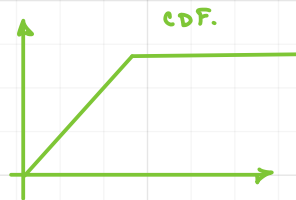
$$\therefore P(X \leq x) = P(F^{-1}(U) \leq x)$$

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

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

$$= F(x)$$

QED.



Ex: Find $n=3$ i.i.d. 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

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

Backsolve: $\therefore x = -\theta \cdot \ln(1-U) \sim \exp(\theta)$

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

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

pick $u_1 = 0.10480 \Rightarrow \therefore x_1 = 0.222$
realization

$u_2 = 0.24130 \Rightarrow \therefore x_2 = 0.552$

$u_3 = 0.22368 \Rightarrow \therefore x_3 = 0.506$

$\therefore$ i.i.d. draws: $\{0.222, 0.552, 0.506\} \sim \exp(2)$



Method: Draw i.i.d. $X_1, \dots, X_n$ 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) $Z_1, \dots, Z_n \sim C(0,1)$

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

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

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

$$U_k = \frac{1}{\pi} \tan^{-1} X_k + \frac{1}{2}$$

$$\therefore Z_k = \tan\left(\pi\left(U_k - \frac{1}{2}\right)\right) \sim C(0,1)$$

$$S \propto S: \alpha \in (0,2)$$

$$\alpha=1 \longrightarrow \text{Cauchy}$$

$$\alpha=2 \longrightarrow \text{Normal}$$

only closed forms.

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

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

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

Generating Normal realizations (Box-Müller, 1958)

Thm: 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

$$X_1 = \sqrt{-2 \ln Y_1} \cos(2\pi Y_2)$$

$$X_2 = \sqrt{-2 \ln Y_1} \sin(2\pi Y_2)$$

Prt 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)

$\therefore$ Can use $f(x_1, x_2) \stackrel{!}{=} |J| \cdot f(y_1, y_2)$ with solutions:

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

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

$\therefore$ The Jacobian is

$$\begin{aligned} J &= \det \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} \\ &= \det \begin{vmatrix} (-x_1) \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) & (-x_2) \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \\ \frac{-x_2/x_1}{2\pi(1+x_2^2/x_1^2)} & \frac{1/x_1}{2\pi(1+x_2^2/x_1^2)} \end{vmatrix} \\ &= \frac{-(1+x_2^2/x_1^2) \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)}{2\pi(1+x_2^2/x_1^2)} = -\frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \end{aligned}$$

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

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

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

$$= f(x_1) \cdot f(x_2) \quad \therefore \text{iid } X_k \sim N(0,1)$$

QED.

Monte Carlo Simulation

"Monte-Carlo" casino in Monaco

2-types

① iid M.C.

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

② "MCMC" Markov Chain MC (EKS17)

- 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

$$\begin{aligned} \textcircled{1} \quad \int_a^b g(x) dx &= \frac{b-a}{b-a} \int_a^b g(x) dx && (a < b) \\ &= (b-a) \int_a^b g(x) \cdot \left(\frac{1}{b-a}\right) g(x) \\ &\stackrel{X \sim U[a,b]}{=} (b-a) \cdot E_x[g(X)]. \end{aligned}$$

② Compute: $Y_k \triangleq (b-a) \cdot g(X_k)$ ↙ simulate
from iid draws $X_k \sim U[a, b]$

$$\begin{aligned} \therefore \bar{Y}_n &= \frac{1}{n} \sum_{k=1}^n Y_k = \frac{(b-a)}{n} \sum_{k=1}^n g(X_k) \\ &\xrightarrow[\text{P}]{\text{LLN}} (b-a) E_x[g(X)] = \int_a^b g(x) dx \\ &\quad \therefore \text{consistent.} \end{aligned}$$

Gaussian Random Vectors

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

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

$n \times 1$ $n \times 1$ $n \times n$ $1 \times n$ $n \times n$ $n \times 1$

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

Defn: X and Y are Jointly Gaussian (JG) iff

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu_z, K_{zz})$$

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

$$\mu_z = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

$$K_{zz} = \begin{pmatrix} K_{xx} & K_{xy} \\ K_{yx} & K_{yy} \end{pmatrix}$$

$\therefore X$ and Y independent if uncorrelated

Defn: If X and Y are JG: the Conditional Gaussian:

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

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

$n \times 1$ $n \times p$ $p \times p$ $p \times 1$

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).

$$\text{iid } X_1, \dots, X_n \sim N(\mu_x, K_{xx})$$

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

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

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)

$$F(A) = \{A, A^c\} \text{ and } P(A) = \{a\}.$$

① Entropy.

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

Unconditional Entropy

$$H(p) = - \sum_{k=1}^n p_k \cdot \log_2 p_k \quad \therefore 0 \leq H(p) \leq \log_2 n.$$

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

Conditional Entropy

$$\begin{aligned} H(Y|X) &= \sum_x p(x) \cdot H(Y|X=x) \\ &= - \sum_x p(x) \sum_y p(y|x) \cdot \log_2 p(y|x) \\ &= - \sum \sum p(x,y) \cdot \log_2 p(y|x) \\ &= - E_{x,y} [\log_2 p(Y|X)] \end{aligned}$$

Mutual Information $I(X,Y)$

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

$$= \sum_x \sum_y p(x,y) \cdot \log_2 \frac{p(x,y)}{p(x)p(y)} \quad \text{kullback-Leibler "divergence"}$$

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

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

$$\begin{aligned}
 &= - \sum_x p(x) \cdot \log_2 p(x) - \left(- \sum_x \sum_y p(x,y) \log_2 p(x|y) \right) \\
 &= H(X) - H(X|Y)
 \end{aligned}$$

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

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

★ Thm: $I(X,Y) \geq 0$ ($= 0 \iff X \text{ and } Y \text{ independent}$)

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

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

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

Jensen's Ineq $\leq \log_2 E_{X,Y} \left[\frac{p(X)p(Y)}{p(X,Y)} \right]$ since logarithm 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.

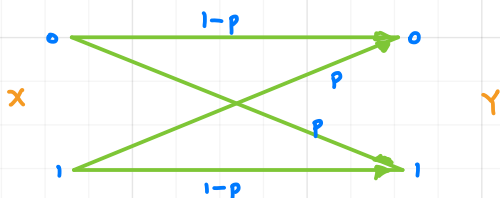
Thm: $H(Y|X) \leq H(Y)$ ★

"Conditioning REDUCES entropy."

Defn: Channel capacity C :

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

Binary Symmetric channel (BSC)



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

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

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

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

$$\uparrow H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

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

$$\leq 1 - H(p)$$

since Y is a binary r.v.

$$\therefore C = 1 - H(p) \text{ bits.}$$

↪ 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 $E_x[X^2] \leq P$

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

Further assume:

- Band-limited $X(t)$ with bandwidth W
- $Y(t) = (X(t) + Z(t)) * h(t)$
 ↑ convolution ← impulse response of ideal bandpass filter. (LTS)

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

for white Gaussian noise with power spectral density $\frac{N_0}{2}$
 $Z \sim N(0, \frac{N_0}{2})$

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

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

← linear function of P .

$\therefore$ infinite bandwidth $\nRightarrow$ infinite capacity

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

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

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

"Typical" sequence

Prf: $-\frac{1}{n} \log_2 p(X_1, \dots, X_n) = -\frac{1}{n} \sum_{k=1}^n \log_2 p(\underbrace{X_k}_{\text{iid}})$

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

$$\begin{aligned} \text{But } -E_X[\log_2 p(X)] &= E_X[\log_2 \frac{1}{p(X)}] \\ &= -\sum_x p(x) \cdot \log_2 p(x) \\ &= H(X) \end{aligned}$$

QED.

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

② Belief Theory

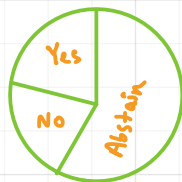
Arthur Dempster (1967)

→ Glen Shafer (1970s)

"Dempster-Shafer" Theory

- allows one to ABSTAIN.

- upper and lower probabilities



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

$$① \quad m(\emptyset) = 0$$

$$② \quad \sum_{A \in 2^X} m(A) = 1 \quad X = \{X_1, \dots, X_n\}$$

$\therefore$ unit mass over 2^X (not sample space X)

$\therefore$ in general:

$$- m(X) < 1$$

$$- m(A^c) \neq 1 - m(A)$$

$$- A \subset B \not\Rightarrow m(A) \leq m(B)$$

$\therefore$ get back discrete probability if all sets are singleton sets

$$\text{Bel}(A) = \text{Belief}(A) = \sum_{B: B \subseteq A} m(B)$$

$\uparrow$ all evidence B that supports A

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

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

$\sim$ all evidence B that fails to refute A .

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

Möbius Inversion Lemma gives:

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

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

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

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

One of uncountably many mass assignments

$$m(H) = 0.4$$

$$m(T) = 0.5$$

$$m(X) = 0.1$$

Total uncertainty in D-S. theory

$$\longleftrightarrow m(X) = 1 \quad (\text{"the keys are somewhere"})$$

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

$$m(X) = 1 \implies m(A) = 0 \text{ if } A \neq X$$

Pick any $A \neq X$

$$\therefore \textcircled{1} \text{ Bel}(A) = 0 \text{ since } m(B) = 0 \text{ for all } B \subseteq A.$$

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

$$\textcircled{2} \text{ Pl}(A) = 1 + 0 + 0 + \dots + 0 = 1 \text{ since } X_n \cap A \neq \emptyset \text{ if } A \neq \emptyset.$$

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

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

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

↑
combined belief

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

Prf:

$$\begin{aligned} 1 &= 1 \cdot 1 = \left(\sum_{B \in 2^X} m_1(B) \right) \left(\sum_{C \in 2^X} m_2(C) \right) \\ &= \sum_{B \in 2^X} \sum_{C \in 2^X} m_1(B) \cdot m_2(C) \\ &= \sum_B \sum_{\substack{C \\ B \cap C = \emptyset}} m_1(B) m_2(C) + \sum_B \sum_{\substack{C \\ B \cap C \neq \emptyset}} m_1(B) \cdot m_2(C) \quad \leftarrow \text{one way to partition} \\ &= \underbrace{\sum_B \sum_{\substack{C \\ B \cap C = \emptyset}} m_1(B) \cdot m_2(C)}_D + \sum_{A \in 2^X} \left(\sum_B \sum_{\substack{C \\ B \cap C = A}} m_1(B) \cdot m_2(C) \right) \end{aligned}$$

QED.

Note: problems arise if D is large

③ Fuzz (Vagueness)

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

$$I_A: X \rightarrow [0,1] \text{ } \leftarrow \text{multivalued range}$$

$$\longleftrightarrow a: X \rightarrow [0,1]$$

$\uparrow$ a is "membership" or "FIT" value
 $a(x) = \text{Degree}(x \in A)$ fuzzy unit.

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

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

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

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

$$A \cup A^c \neq X$$

$$\text{midpoint set } A \longleftrightarrow a(x) \stackrel{\Delta}{=} \frac{1}{2}$$

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

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

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

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

$\uparrow$ Boolean n -cube

Ex: $X = \{x_1, x_2\}$

$$A = (a_1, a_2) = \left(\frac{1}{3}, \frac{3}{4}\right)$$

$$x = \begin{matrix} x_1 & x_2 \\ (1, 1) \end{matrix}$$

$$\phi = (0, 0)$$

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

← "list vector"

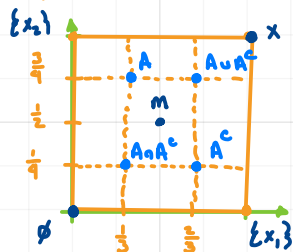
$$A^c = \left(\frac{2}{3}, \frac{1}{4}\right)$$

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

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

$$m = \left(\frac{1}{2}, \frac{1}{2}\right)$$

"Sets as points"
 $A \in [0, 1]^2$



Fuzzy Counting Measure $c(A)$:

$$c(A) = \begin{cases} \sum_x a(x) \\ \int_x a(x) dx \end{cases}$$

discrete case

continuous case

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

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

not an integer in general.

Fuzzy Equality:

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

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

$$F(m) = 1$$

$$F(B) = 0 \text{ if } B \text{ binary}$$

Subsethood (Fuzzy Inclusion) S :

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

↑
resembles $P(B|A)$

$$\therefore F(A) = S(A \cup A^c, A \cap A^c)$$

"whole" "part" $\therefore$ "whole in part"



Thrm:

$F(A) = S(A \cup A^c, A \cap A^c) = \mathcal{E}(A, A^c)$

↑
Fuzziness

↑
Subsethood

↑
Equality

Ex: Suppose $\exists$ probability measure P such that $P \stackrel{\Delta}{=} F$
(i.e. here is "disguised" probability)

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

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

$\therefore 0 < 0$ → ← contradiction

$\therefore$ There is no such P .

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

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

$\therefore$ 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

$\therefore$ reduces to subsethood

Probability of a Fuzzy Event. (Zadeh, 1968)

Suppose $a: X \rightarrow [0, 1]$ is Borel-measurable

$$\therefore P(A) = E[a]$$

$$(P(A) = E_x[I_A])$$

$$= \sum_x p(x) a(x)$$

$$\left(= \int_x f(x) a(x) dx < \infty \right)$$

$\therefore$ In general: $P(A \cap A^c) > 0$

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

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

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

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

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

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

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

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