ECSE 6810 PGM HW.

Shiuli Subhra Ghosh.
RIN - 662056315

1. We wish to calculate $P(H, E_1, E_2)$

(b) $P(E_1, E_2)$, $P(H)$ and $P(E_1 E_2|H)$ is sufficient.

Because

$P(H, E_1, E_2) = P(H|E_1,E_2) \cdot P(E_1, E_2) = \underline{P(E_1, E_2|H) \cdot P(H)}$

So, From the conditional Expansion, $P(E_1, E_2|H) \cdot P(H)$ is sufficient we don't even need to calculate $P(E_1 E_2)$.

however (a) $P(E_1,E_2)$, $P(H)$, $P(E_1|H)$ is and $P(E_2|H)$ this is insufficient. In particular we can't obtain $P(E_1, E_2|H)$ from $P(E_1|H)$ and $P(E_2|H)$ unless we know that E_1 and E_2 are independent. Knowing marginal distribution of $P(E_1 E_2)$ will not help in this case.

(c) $P(H)$, $P(E_1|H)$ and $P(E_2|H)$ is also not sufficient because it is same as a ~~b~~ only without $P(E_1, E_2)$. In fact we don't need to know $P(E_1, E_2)$. We can find it using the conditional sum $= P(E_1, E_2) = \sum_{H=a} P(H=a) \times P(E_1 E_2|H=a)$.
If we know $P(E_1, E_2|H)$. But again it is impossible to find $P(E_1, E_2|H)$ from $P(E_1|H)$ and $P(E_2|H)$ unless we know that E_1 and E_2 are independent.

Now if we know E_1 and E_2 are independent.

$$P(H, E_1, E_2) = P(H) \cdot P(E_1, E_2 | H).$$
$$= P(H) \cdot P(E_1|H) \cdot P(E_2|H).$$

So,
- $P(E_1, E_2)$, $P(H)$, $P(E_1|H)$, $P(E_2|H)$
- $P(E_1, E_2)$, $P(H)$, $P(E_1, E_2|H)$
- $P(E_1|H)$, $P(E_2|H)$ and $P(H)$

all 3 are sufficient.

2. (a) $X \perp Y, W | Z \Rightarrow X \perp Y | Z$ (true)

$X \perp Y, W | Z \Rightarrow P(X, Y, W | Z) = P(X|Z) \cdot P(Y, W|Z).$

lets marginalize $P(X, Y, W | Z)$ on W.

$$\sum_{W} P(X, Y, W | Z) = \sum_{W} P(X|Z) \cdot P(Y, W|Z)$$
$$= P(X|Z) \sum_{W} P(Y, W|Z)$$
$$= P(X|Z) \sum_{W} P(W|Z) \cdot P(Y|W, Z)$$
$$= P(X|Z) \cdot P(Y|Z)$$

If there are 3 RV's X, Y, Z and if X, Y are independent given Z, then $P(X, Y|Z) = P(X|Z) \cdot P(Y|Z)$

and we denote it as $X \perp Y | Z$

So, $P(X|Z) \cdot P(Y|Z) \Rightarrow X \perp Y | Z$. hence the identity is true.

(b) $X \perp Y | Z$ and $X, Y \perp W | Z \Rightarrow X \perp W | Z$.

$X \perp Y | Z \Rightarrow P(X, Y | Z) = P(X | Z) . P(Y | Z)$.

$X, Y \perp W | Z \Rightarrow P(X, Y, W | Z) = P(X, Y | Z) . P(W | Z)$

$= P(X | Z) . P(Y | Z) . P(W | Z)$

So, X, Y, Z, W are 4 RVs. out of which X, Y and W are independent given Z.

marginalize on Y,

$\sum_y P(X, Y, W | Z) = P(X, W | Z)$.

or, $\sum_y P(X | Z) . P(Y | Z) . P(W | Z) = P(X, W | Z)$

or, $P(X | Z) . P(W | Z) \sum_y P(Y | Z) = P(X, W | Z)$

or, $P(X | Z) . P(W | Z) . 1 = P(X, W | Z)$.

hence $X \perp W | Z$. (proved).

(c). $X \perp Y, W \mid Z$ and $Y \perp W \mid Z$.

So, $X \perp Y, W \mid Z \Rightarrow P(X, Y, W \mid Z) = P(X \mid Z) . P(Y, W \mid Z)$

$Y \perp W \mid Z \Rightarrow P(Y, W \mid Z) = P(Y \mid Z) . P(W \mid Z)$.

hence, $P(X, Y, W \mid Z) = P(X \mid Z) . P(Y \mid Z) . P(W \mid Z)$

$= \cancel{P}$

lets marginalize w.r.t Y.

So, $\sum_y P(X, Y, W \mid Z) = P(X, W \mid Z) . = \sum_y P(X \mid Z) P(Y \mid Z) P(W \mid Z)$.

$= P(X \mid Z) . P(W \mid Z) . 1$.

So, $X \perp W \mid Z$

So, $P(X, Y, W \mid Z) = P(X \mid Z) . P(Y \mid Z) . P(W \mid Z)$

$= P(X, W \mid Z) . P(Y \mid Z)$

hence from this factorization we get.

$X, W \perp Y \mid Z$. (proved)

(d) $X \perp Y \mid Z$ and $X \perp Y \mid W \Rightarrow X \perp Y \mid Z, W$ (False)

Counter example –

Let X, Y, Z ~~and W~~ are iid , each with equal probabilities of being -1 and 1.

Let $W = XYZ$

So, in this example , $X \perp Y \mid Z$

$X \perp Y \mid W$

But when we consider 4 variables, and if we know Z and W then the independence doesn't hold true. because we know $W = XYZ$.

Hence $X \not\perp Y \mid Z, W$.

Also, graphical example (done a course on Causality (ETH Zurich))

$\cancel{X \rightarrow Z \rightarrow D \leftarrow W}$. $X \rightarrow Z \leftarrow D \rightarrow W \leftarrow Y$

there is a path from X to Y. Z and W are the colliders. So, If we don't condition on both the colliders the path is blocked. Because there is a confounder D. But conditioning on both Z and W makes the path Active.

So, $X \perp Y \mid Z$

$X \perp Y \mid W$.

$X \not\perp Y \mid Z, W$.

hence the result is False.

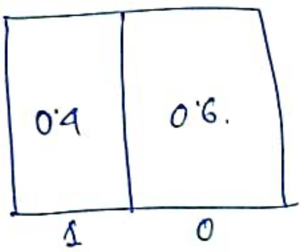
3. 1) Given a binary random variable $X \in \{0,1\}$ with $P[X=1] = 0.4$ we need to draw N random samples where $N = [50, 100, 150, 200]$.

So, we will follow Monte carlo simulation for random sample generation.

here the random variable can take 2 values.

Hence we will divide the region between 0 and 1 in 2 parts using parameter θ.

here $\theta = 0.4$.



then we will generate $N = [50, 100, 150, 200]$ random variables following $U(0,1)$.

$X = 1$ if u is in between ~~0.6 to~~ $[0, 0.4]$
0 to 0.4.

$X = 0$ if u is in between 0.4 to 1.
$(0.4, 1]$

So, this is the sampling method we can follow to generate N numbers of samples.

Given the samples, now we will obtain the MLE of $P[X=1]$, which is $= \dfrac{\text{number of times } x=1}{N}$

4. For a Bernoulli trial involving coin toss, assuming that we have tossed the coin N times resulting a Data

$D = [x_1, x_2, \ldots, x_N]$

out of which we observed head up $(x=1)$ k times.

Let θ be the probability of head up.

$$\theta = P(x=1)$$

Given D, we need to find θ_{MLE}, θ_{MAP} and θ_{Bay} of θ.

<u>Derivation of θ_{MLE}</u>

$$\theta_{MLE} = \arg\max_{\theta} \log P(D|\theta) \qquad x_1, x_2 \ldots, x_N \overset{iid}{\sim} ber(\theta)$$

$$\log P(D|\theta) = \log P(x_1, x_2 \ldots, x_N | \theta)$$

$$= \log \prod_{i=1}^{N} P(x_i|\theta)$$

$$= \sum_{i=1}^{N} \log P(x_i|\theta)$$

$$= \sum_{i=1}^{N} \log \theta^{\mathbb{I}(x_i=1)} (1-\theta)^{\mathbb{I}(x_i=0)}$$

$$= \sum_{i=1}^{N} \mathbb{I}(x_i=1) \log \theta + \sum_{i=1}^{N} \mathbb{I}(x_i=0) \log(1-\theta)$$

$$= n \log \theta + (N-n) \log(1-\theta)$$

n = no. of times head occurs.

Given Log-likelihood take the first order derivative w.r.t θ

$$\frac{d}{d\theta} \log P(D|\theta) = \frac{n}{\theta} + \frac{N-n}{\theta-1} = 0$$

or, $\frac{n}{\theta} = \frac{N-n}{1-\theta}$

or, $\frac{1-\theta}{\theta} = \frac{N-n}{n}$. So, $\hat{\theta}_{MLE} = \frac{n}{N}$.

or, $\frac{1}{\theta} = \frac{N}{n}$

or, $\theta = \frac{n}{N}$

<u>Derivation of θ_{MAP}</u>

$$\theta_{MAP} = \arg\max_{\theta} \log P(\theta|D)$$

$$= \arg\max_{\theta} \left[\log P(\theta) + \log P(D|\theta)\right]$$

$$= \arg\max_{\theta} \left\{\log P(\theta) + \sum_{i=1}^{N} \log P(D_i|\theta)\right\}$$

Conjugate prior of Binomial distribution is Beta distribution.

So, $\theta \sim Beta(\alpha, \beta)$

$$\theta_{MAP} = \arg\max_{\theta} \left\{ \log \frac{\Gamma\alpha \, \Gamma\beta.}{\sqrt{\alpha+\beta}} + (\alpha+1)\log\theta + (\beta-1)\log(1-\theta) + n\log\theta + (N-n)\log(1-\theta) \right\}.$$

derivative w.r.t θ.

$$\frac{\alpha-1}{\theta} - \frac{\beta-1}{1-\theta} + \frac{n}{\theta} - \frac{N-n}{1-\theta} = 0.$$

Bayesian Estimator Derivation :-

$\theta \sim \text{Beta}(\alpha, \beta)$

hence

$P(\theta|D) \propto \theta^{n}(1-\theta)^{N-n} \quad \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$n =$ no. of success.

So, $P(\theta|D) \propto \theta^{n+\alpha-1}(1-\theta)^{N-n+\beta-1}$

to make the density a new beta density,

$P(\theta|D) = c \cdot \theta^{n+\alpha-1}(1-\theta)^{N-n+\beta-1}$

we need to choose c such that $\int_{-\infty}^{\infty} P(\theta|D)\, d\theta = 1$

So, as $P(\theta|D)$ follows beta with $(n+\alpha)$ and $(N-n+\beta)$

then $c = \dfrac{\Gamma(n+\alpha+\beta)}{\Gamma(n+\alpha)\,\Gamma(N-n+\beta)}$

So, to find the Bayesian estimate, we need to find the mean of Beta distribution $(n+\alpha, N-n+\beta)$.

So, mean of Beta $= \dfrac{n+\alpha}{n+\alpha+N-n+\beta} = \boxed{\dfrac{n+\alpha}{N+\alpha+\beta} = \hat{\theta}_{Bay}}$

If $\alpha = \beta = 1$ then.

$\hat{\theta}_{Bay} = \dfrac{n+1}{N+2}.$

$$\frac{\alpha-1}{\theta}+\frac{n}{\theta}=\frac{\beta-1}{1-\theta}+\frac{N-n}{1-\theta}$$

or, $$\frac{\alpha+n-1}{\theta}=\frac{\beta+N-n-1}{1-\theta}$$

or, $$\frac{1-\theta}{\theta}=\frac{\beta+N-n-1}{\alpha+n-1}$$

or, $$\frac{1}{\theta}=\frac{\beta+N-\cancel{n}-1+\alpha+\cancel{n}-1}{\alpha+n-1}$$

or, $$\frac{1}{\theta}=\frac{\beta+N+\alpha-2}{\alpha+n-1}$$

or, $$\theta=\frac{\alpha+n-1}{\beta+N+\alpha-2}$$

$$\hat{\theta}_{MAP}=\frac{\alpha+n-1}{\beta+N+\alpha-2}$$

If $\theta \sim \text{Beta}(1,1)$ then $\hat{\theta}_{MAP}=\frac{n}{N}$.

4.(b) As all the tosses are independent, the probability of $N+1$ th toss of being 1 is same as θ.

$$P[X_{N+1} = 1] = \theta.$$

hence $\hat{\theta}_{MLE} = \dfrac{n}{N}$ $= P[X_{N+1} = 1 | \hat{\theta}_{MLE}]$

$\hat{\theta}_{MAP} = \dfrac{n}{N}$ [Assuming prior. dist of θ follows Beta(1,1)]

$\hat{\theta}_{Bay} = \dfrac{n+1}{N+2}$. ''

4.(c) Now we need to perform a Bayesian prediction of the probability that head is up for the $N+1$ th toss

$$P[X_{N+1} = 1 | D]$$

$$= \sum_{\theta} P[X_{N+1} = 1, \theta | D]$$

$P[X_{N+1} = 1 | D, \theta]$
$= P[X_{N+1} = 1 | \theta]$

$$= \int_{\theta} P(\theta | D) . P(X_{N+1} | D, \theta) \, d\theta$$

[Because once we know the prior estimate, we don't need to ~~calculate~~ consider the sample till N]

$$= \int_{\theta} P(\theta | D) . P(X_{N+1} = 1 | \theta) \, d\theta$$

$$= \int_{\theta} \theta \, P(\theta | D) \, d\theta = E_{P(\theta|D)}(\theta)$$

$$= \frac{n+\alpha}{N+\alpha+\beta}.$$

$$= \frac{n+1}{N+2} \quad [\text{if } \alpha = \beta = 1]$$