Stats305B Problem Set 1

1 Agresti 1.29

1.1 a.)

1.29a $L(\theta) \propto \theta^{2n_1}[\theta(1-\theta)]^{n_2}(1-\theta)^{2\underbrace{(n-n_1-n_2)}_{n_3}}$

$$\Longrightarrow \underbrace{\ell(\theta)}_{\text{KERNEL}} = 2n_1 \log\theta + n_2\log(\theta) + n_2\log(1-\theta)$$
$$+ 2n\log(1-\theta) - 2n_1\log(1-\theta) - 2n_2\log(1-\theta)$$
$$= [2n_1+n_2]\log\theta + [2n-2n_1-n_2]\log(1-\theta)$$

$$\Longrightarrow \ell'(\theta) = \frac{2n_1+n_2}{\theta} - \frac{2n-2n_1-n_2}{(1-\theta)}$$

$$\Longrightarrow \ell'(\theta)=0 \Longrightarrow \frac{2n_1+n_2}{\theta} = \frac{2n-2n_1-n_2}{(1-\theta)}$$

$$\hookrightarrow 2n_1+n_2-\cancel{2n_1\theta}-\cancel{n_2\theta} = 2n\theta-\cancel{2n_1\theta}-\cancel{n_2\theta}$$
$$2n_1+n_2 = 2n\theta$$

$$\Longrightarrow \boxed{\hat{\theta} = \frac{2n_1+n_2}{n}},$$ as desired ∎

Note $\ell''(\theta)$ is convex (see below), ensuring $\ell'(\theta)=0$ is critical point

1.29b So, by chain rule and negation

$$\ell''(\theta) = -\frac{2n_1+n_2}{\theta^2} - \frac{2n-2n_1-n_2}{(1-\theta)^2}$$

$$\Rightarrow -\ell''(\theta) = \frac{2n_1+n_2}{\theta^2} + \frac{2n-2n_1-n_2}{(1-\theta)^2}$$

$$= \frac{2n_1+n_2}{\theta^2} + \frac{\cancel{2n_1}+2n_2+2n_3-\cancel{2n_1}-n_2}{(1-\theta)^2}$$

$$= \boxed{\frac{2n_1+n_2}{\theta^2} + \frac{2n_3+n_2}{(1-\theta)^2}}$$ ∎

1.2 b.)

$$E[\ell''(\theta)] = \frac{1}{\theta^2} \underbrace{E[2n_1 + n_2]}_{\text{PLUG IN}} + \frac{1}{(1-\theta)^2} E[2n_3 + n_2]$$

$$= \frac{1}{\theta^2} 2n\theta^2 + \frac{1}{\theta^2} 2n\theta(1-\theta) + \frac{1}{(1-\theta)^2} 2n(1-\theta)^2 + \frac{1}{(1-\theta)^2} 2n(\theta)(1-\theta)$$

$$= 2n + 2n + \frac{2n(1-\theta)}{\theta} + \frac{2n\theta}{(1-\theta)}$$

$$= \frac{4n\theta(1-\theta)}{\theta(1-\theta)} + \frac{2n(1-\theta)^2}{\theta(1-\theta)} + \frac{2n\theta^2}{\theta(1-\theta)}$$

$$= \frac{4n\theta - 4n\theta^2 + 2n - 4n\theta + 2n\theta^2 + 2n\theta^2}{\theta(1-\theta)}$$

$$= \boxed{\frac{2n}{\theta(1-\theta)}}$$

Since $\lim_{n\to\infty} Var(\theta_n) \to I(\theta_n)^{-1}$ as $n \to \infty$,

we have as $n \to \infty$

$$Var(\theta_n) \to \frac{\theta(1-\theta)}{2n}$$

so $\Longrightarrow$ $\boxed{SE(\hat{\theta}_n) \to \left(\frac{\hat{\theta}(1-\hat{\theta})}{2n}\right)^{1/2}}$ ← estimated

1.3 c./d.)

1.29d Setup your standard χ^2 test, with

- $E[n_1] = n\hat{p}_1 = n\hat{\theta}^2$
- $E[n_2] = n\hat{p}_2 = 2n\hat{\theta}(1-\hat{\theta})$
- $E[n_3] = n\hat{p}_3 = n(1-\hat{\theta})^2$
- $df = (3-1) - 1 = 1$

n_3 is deterministic from n, n_1, n_2: 3 params, one superfluous

$\Rightarrow$ Test

$$\tau = (n_1 - n\hat{\theta}^2)^2 + (n_2 - 2n\hat{\theta}(1-\hat{\theta}))^2 + (n_3 - (1-\hat{\theta})^2 n)^2$$

vs χ^2_1, reject for large τ.

H_0 null hypothesis, $df = 3-1-1 = 1$

2 Agresti 2.27

2.1 Interpretation

First, let's interpret the numerator. If we take $P(D)$ to represent your overall probability of contracting the disease (exposure or no exposure), and $P(D|E')$ to be the probability that you'd have gotten the disease anyways even if completely unexposed (i.e. living in quarantine, but still somehow got it), then the difference between the two provides an approximate delta of how much exposure is at fault vs. the counterfactual of staying at home – i.e., how much of the total frequency we can attribute to the exposure. Finally, division by $P(D)$ provides the standardization, converting this delta into an attributable proportion among the diseased.

2.2 Proof

Working right-to-left, we have, using $RR = \frac{P(D|E)}{P(D|E')}$

$$
\begin{aligned}
\frac{P(E)(RR-1)}{1+P(E)(RR-1)} &= \frac{P(E)(\frac{P(D|E)}{P(D|E')}-1)}{1+P(E)(\frac{P(D|E)}{P(D|E')}-1)} \\
&= \frac{P(E)(\frac{P(D|E)}{P(D|E')}-\frac{P(D|E')}{P(D|E')})}{\frac{P(D|E')}{P(D|E')}+P(E)(\frac{P(D|E)}{P(D|E')}-\frac{P(D|E')}{P(D|E')})} \\
&= \frac{\left(\frac{1}{P(D|E')}\right)\Big((P(E)P(D|E)-P(E)P(D|E')\Big)}{\left(\frac{1}{P(D|E')}\right)\Big((P(D|E')+P(E)P(D|E)-P(E)P(D|E'))\Big)} \\
&= \frac{(P(E)P(D|E)-P(E)P(D|E')}{(P(D|E')+P(E)P(D|E)-P(E)P(D|E'))} \\
&= \frac{P(E)P(D|E)-(1-P(E'))P(D|E')}{P(D|E')+P(E)P(D|E)-(1-P(E'))P(D|E')} \\
&= \frac{P(E)P(D|E)-P(D|E')+P(E')P(D|E')}{P(E)P(D|E)+P(D|E')-P(D|E')+P(E')P(D|E')} \\
&= \frac{P(E)P(D|E)+P(E')P(D|E')-P(D|E')}{P(E)P(D|E)+P(E')P(D|E')} \\
&= \frac{\overbrace{P(E)P(D|E)+P(E')P(D|E')}^{P(D);\ LoTP}-P(D|E')}{\underbrace{P(E)P(D|E)+P(E')P(D|E')}_{P(D);\ LoTP}} \\
&= \frac{P(D)-P(D|E')}{P(D)},
\end{aligned}
$$

as desired.

3 Matching/Case Control

3.1 a.

First,we take a SRS from the population of those with cancer (C=1), giving us some observed estimate $\hat{P}(A,S|C=1) \sim P(A,S|C=1)$. Then, among the population that does not have cancer, we sample and match those $C=0$ samples to each case in the $C=1$ cancer group, matching/joining on age and sex. Hence, if we sample indefinitely, we should expect $\hat{P}(A,S|C=1)$ to settle down to $P(A,S|C=1)$, and with the matching/pairing scheme enforcing the same sample proportions for the $C=2$ group, the distribution of age and sex within $C=2$ should also settle to $P(A,S|C=1)$.

3.2 b.

Given $N \sim Pois(\lambda)$, we know from lecture that for $N = n$,

$$N_{ij}|N = n \sim Binom(n, p_{ij}).$$

The goal here is to obtain a density for N_{ij} unconditional, so we margin out/sum over N, i.e.

$$\begin{aligned}
P(N_{ij} = n_{ij}) &= \sum_{\ell=0}^{\infty} \underbrace{P(N_{ij} = n_{ij}|N = \ell)}_{known\ binomial} \underbrace{P(N = \ell)}_{original\ Poisson} \\
&= \sum_{\ell=0}^{\infty} \binom{\ell}{n_{ij}} p_{ij}^{n_{ij}} (1-p_{ij})^{\ell - n_{ij}} \left(\frac{\exp(-\lambda)\lambda^{\ell}}{\ell!}\right) \\
&= \sum_{\ell=n_{ij}}^{\infty} \frac{\ell!}{n_{ij}!(\ell - n_{ij})!\ell!} p_{ij}^{n_{ij}} (1-p_{ij})^{\ell - n_{ij}} \exp(-\lambda)\lambda^{\ell} \\
&= \exp(-\lambda)\left(\frac{p_{ij}^{n_{ij}}}{n_{ij}!}\right) \sum_{\ell=n_{ij}}^{\infty} \left(\frac{(1-p_{ij})^{\ell-n_{ij}}}{(\ell - n_{ij})!}\lambda^{\ell}\right) \\
&= \exp(-\lambda)\left(\frac{p_{ij}^{n_{ij}}}{n_{ij}!}\right) \sum_{\ell=n_{ij}}^{\infty} \left(\frac{(1-p_{ij})^{\ell-n_{ij}}}{(\ell - n_{ij})!}\lambda^{n_{ij}-n_{ij}+\ell}\right) \\
&= \exp(-\lambda)\left(\frac{p_{ij}^{n_{ij}}}{n_{ij}!}\right)\lambda^{n_{ij}} \sum_{\ell=n_{ij}}^{\infty} \left(\frac{(1-p_{ij})^{\ell-n_{ij}}}{(\ell - n_{ij})!}\lambda^{\ell-n_{ij}}\right) \\
&= \exp(-\lambda)\left(\frac{(\lambda p_{ij})^{n_{ij}}}{n_{ij}!}\right) \sum_{\ell=n_{ij}}^{\infty} \left(\frac{(\lambda(1-p_{ij}))^{\ell-n_{ij}}}{(\ell - n_{ij})!}\right) \\
&= \exp(-\lambda)\left(\frac{(\lambda p_{ij})^{n_{ij}}}{n_{ij}!}\right) \sum_{k=0}^{\infty} \left(\frac{(\lambda(1-p_{ij}))^{k}}{k!}\right).
\end{aligned}$$

We pause to recall the Taylor expansion

$$\exp(x) = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$\Longrightarrow$$

$$\exp(\lambda(1-p_{ij})) = \sum_{k=0}^{\infty} \frac{(\lambda(1-p_{ij}))^k}{k!},$$

so we see

$$\begin{aligned}
P(N_{ij} = n_{ij}) &= \exp(-\lambda)\left(\frac{(\lambda p_{ij})^{n_{ij}}}{n_{ij}!}\right) \sum_{k=0}^{\infty} \left(\frac{(\lambda(1-p_{ij}))^{k}}{k!}\right) \\
&= \exp(-\lambda)\left(\frac{(\lambda p_{ij})^{n_{ij}}}{n_{ij}!}\right) \exp(\lambda(1-p_{ij})) \\
&= \left(\frac{(\lambda p_{ij})^{n_{ij}}}{n_{ij}!}\right) \exp(-\lambda + \lambda(1-p_{ij})) \\
&= \left(\frac{(\lambda p_{ij})^{n_{ij}}}{n_{ij}!}\right) \exp(\lambda p_{ij})) \\
&\sim Pois(\lambda p_{ij}),
\end{aligned}$$

which is an intuitive result. Applying this result to each cell thus gives distributions of:

$$N \sim Pois(\lambda);\ N_{11}|N \sim Binom(n, p_{11}) \implies N_{11} \sim Pois(\lambda p_{11}) = Pois(\lambda P(T = 1|C = 1))$$

and identically

$$N_{21} \sim Pois(\lambda p_{21}) = Pois(\lambda P(T = 2|C = 1)).$$

Due to the SRS nature of the cancer C=1 draws here, this computation is straightforward, and we may compute $N_{i1} \sim Pois(\lambda p_{i1})$ directly "from" $P(T = i|C = 1)$.

By contrast, the matched/fixed/engineered nature of the $C = 2$ non-cancer control group requires that we incorporate the age A and sex S random variables by marginalizing them out, i.e.

$$\begin{aligned} P(T = 1|C = 2) &= \sum_s \sum_a P(T = 1|S = s, A = a, C = 2)P(S = s, A = a|C = 2) \\ &\approx \sum_s \sum_a P(T = 1|S = s, A = a, C = 2)\, \underbrace{P(S = s, A = a|C = 1)}_{\textit{by matching}}, \end{aligned}$$

where the $P(S = s, A = a|C = 2) \approx P(S = s, A = a|C = 1)$ comes by construction/from the matching procedure, wherein we strategically sampled $P(S = s, A = a|C = 2)$ to resemble the observed $P(S = s, A = a|C = 1)$. Hence, we must use (presenting three equivalent definitions here, in case there's some sequential sampling scheme at play that makes one ordering more intuitive):

$$\begin{aligned} p_{j2} &= \sum_{s,a} P(T = j|S = s, A = a, C = 2)P(S = s, A = a|C = 1) \\ &= \sum_{s,a} P(T = 1|S = s, A = a, C = 2)P(S = s|A = a, C = 1)P(A = a|C = 1) \\ &= \sum_{s,a} P(T = 1|S = s, A = a, C = 2)P(A = a|S = s, C = 1)P(S = s|C = 1). \end{aligned}$$

Then, from the result above, we have

$$N_{12} \sim Pois(\lambda p_{12})$$

and

$$N_{22} \sim Pois(\lambda p_{22}).$$

3.3 c.

From our findings for N_{ij} above, it follows that

$$\begin{aligned} OR &= \frac{N_{11}N_{22}}{N_{12}N_{21}} \\ &\xrightarrow{P} \frac{\lambda^2 P(T = 1|C = 1) \sum_s \sum_a P(T = 2|S = s, A = a, C = 2)\, \underbrace{P(S = s, A = a|C = 1)}_{\textit{matching}}}{\lambda^2 P(T = 2|C = 1) \sum_s \sum_a P(T = 1|S = s, A = a, C = 2)\, \underbrace{P(S = s, A = a|C = 1)}_{\textit{matching}}} \\ &= \frac{P(T = 1|C = 1) \sum_s \sum_a P(T = 2|S = s, A = a, C = 2)P(S = s, A = a|C = 1)}{P(T = 2|C = 1) \sum_s \sum_a P(T = 1|S = s, A = a, C = 2)P(S = s, A = a|C = 1)}. \end{aligned}$$

However, due to the matched construction of the control group (as based on the cancer group), sex and age here are effectively fixed: hence, within the $\sum_{s,a}$, the only s, a for which $P(S = s, A = a|C = 1)$ does not zero out are those s, a corresponding to the sample's matching configuration, and hence we have

$$\begin{aligned}
OR &= \frac{P(T=1|C=1)\sum_s \sum_a P(T=2|S=s, A=a, C=2)P(S=s, A=a|C=1)}{P(T=2|C=1)\sum_s \sum_a P(T=1|S=s, A=a, C=2)P(S=s, A=a|C=1)} \\
&= \frac{P(T=1|C=1)P(T=2|S=s, A=a, C=2)P(S=s, A=a|C=1)}{P(T=2|C=1)P(T=1|\underbrace{S=s, A=a}_{\text{“}fixed\ under\ matching\text{''}}, C=2)P(\underbrace{S=s, A=a}_{\text{“}fixed\ under\ matching\text{''}}|C=1)} \\
&= \frac{P(T=1|C=1)P(T=2|S=s, A=a, C=2)}{P(T=2|C=1)P(T=1|S=s, A=a, C=2)}.
\end{aligned}$$

Observing that the "true" odds ratio can be expressed as (see. Agresti 2.24 Eqn. 2.6)

$$OR = \frac{P(T=1, C=1)P(T=2, C=2)}{P(T=2, C=1)P(T=1, C=2)} = \frac{P(T=1|C=1)P(T=2|C=2)}{P(T=2|C=1)P(T=1|C=2)}$$

we now see that the above expression looks extremely similar – with the single caveat that when taking the $P(T=t|C=2)$ cases, we must also condition on age and sex, in order to bake the matching procedure into any sample odds ratio here.

4 Pearson and LRT

Here, we recreate the proof on page 596-597. For notation, we let $p_{ij} = n_{ij}/N$ be the sample proportions in each cell, $\hat{\pi}_{ij}$ be the "model-based" estimator (i.e. that by taking the column-wise and row-wise proportions/MLEs, and multiplying under the assumption of independence), let $\hat{\mu}_{ij} = n\hat{\pi}_{ij}$ and let π_{ij} be the true distribution. We then have for G^2

$$\begin{aligned}
G &= 2\sum_{i,j} n_{ij} \log\left(\frac{n_{ij}}{\hat{u}_{ij}}\right) \\
&= 2\sum_{i,j} n_{ij} \log\left(\frac{Np_{ij}}{N\hat{\pi}_{ij}}\right) \\
&= 2\sum_{i,j} n_{ij} \log\left(\frac{p_{ij}}{\hat{\pi}_{ij}}\right) \\
&= 2\sum_{i,j} n_{ij} \log\left(\frac{p_{ij} - \hat{\pi}_{ij} + \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) \\
&= 2\sum_{i,j} n_{ij} \log\left(1 + \frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) \\
&= 2N\sum_{i,j} p_{ij} \log\left(1 + \frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right).
\end{aligned}$$

At this point, stop to recall that a Taylor series expansion gives

$$\log(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k,$$

so

$$\log\left(1 + \frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}\left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right)^k,$$

and thus

$$
\begin{aligned}
G &= 2N \sum_{i,j} p_{ij} \log\left(1 + \frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) \\
&= 2N \sum_{i,j} p_{ij} \Bigg[\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right)^k \\
&= 2N \sum_{i,j} [\hat{\pi}_{ij} + (p_{ij} - \hat{\pi}_{ij})] \left[\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right)^k \right] \\
&= 2N \sum_{i,j} [\hat{\pi}_{ij} + (p_{ij} - \hat{\pi}_{ij})] \left[\left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) - \frac{1}{2}\left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right)^2 + \dots \right] \\
&= 2N \sum_{i,j} \left[\hat{\pi}_{ij} \left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) - \frac{1}{2}\left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right)^2 + \dots \right] \\
&\qquad\qquad + \left[(p_{ij} - \hat{\pi}_{ij}) \left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right) - \frac{1}{2}\left(\frac{p_{ij} - \hat{\pi}_{ij}}{\hat{\pi}_{ij}}\right)^2 + \dots \right] \\
&= \underbrace{2N \sum_{ij} (p_{ij} - \hat{\pi}_{ij})}_{\sum_{i,j} p_{ij} = \sum_{i,j} \hat{\pi}_{ij} \implies 0} - \frac{1}{2}\left(\frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}\right) + O_p^{(1)}(p_{ij} - \hat{\pi}_{ij})^3 + \frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}} + O_p^{(2)}(p_{ij} - \hat{\pi}_{ij})^3 \\
&= 2N \sum_{ij} \frac{1}{2}\left(\frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}\right) + O_p(p_{ij} - \hat{\pi}_{ij})^3 \\
&= N \sum_{ij} \left(\frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}\right) + 2N O_p(p_{ij} - \hat{\pi}_{ij})^3 \\
&= N \sum_{ij} \left(\frac{(p_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}\right) + 2N O_p(n^{3/2}) \\
&= X^2 + O_p(n^{-1/2}) \\
&= X^2 + o_p(1),
\end{aligned}
$$

which tells us that $X^2 - G^2 \xrightarrow{P} 0$, as desired. As set forth in Agresti, the sequence of $O_p(p_{ij} - \hat{\pi}_{ij})^3 \to O_p(n^{3/2})$ works as follows. First, we have by straightforward expansion that

$$
\begin{aligned}
O_p([]p_{ij} - \hat{\pi}_{ij}]^3) &= O_p([\underbrace{(p_{ij} - \pi_{ij})}_{\xi_1} - \underbrace{(\hat{\pi}_{ij} - \pi_{ij})}_{\xi_2}]^3) \\
&= O_p((\xi_1 - \xi_2)^3) \\
&= O_p(\xi_1^3 - 3\xi_1^2\xi_2 + 3\xi_1\xi_2^2 - \xi_2^3).).
\end{aligned}
$$

Then, since it is known that both ξ_1, ξ_2 are $O_p(n^{-1/2})$, we have that each of $\xi_1^3, \xi_1^2\xi_2, \xi_1\xi_2^2, \xi_2^3$ are each $O_p((n^{-1/2})^3) = O_p(n^{-3/2})$. That is,

$$
\begin{aligned}
O_p(\xi_1^3 - 3\xi_1^2\xi_2 + 3\xi_1\xi_2^2 - \xi_2^3).) &= O_p((n^{-1/2})^3) \\
&= O_p(n^{-3/2}).
\end{aligned}
$$

This gives the first step. Then, converting to little-o notation, know from 16.1.1 that $O_p(n^{-1/2})$ is $o_p(1)$, and since $O_p(n^{-3/2})$ gets to zero faster than does $O_p(n^{-1/2})$, it is also the case that $O_p(n^{-3/2})$ is $o_p(1)$. And as set forth in the final paragraph of 16.1.1, a difference of $o_p(1)$ between two random variables (here, G^2, X^2 means that as $n \to \infty$, those two random variables achieve the same limiting distribution; in this context, this means that X^2 and G^2 have identical limiting distributions, and hence $X^2 - G^2 \xrightarrow{P} 0$, as desired.

5 Homogeneous Association

5.1

As given by the problem, we have

$$P(X=x,Y=y,Z=z) \propto \exp(\alpha_x x + \alpha_y y + \alpha_z z + \alpha_{xy}xy + \alpha_{xz}xz + \alpha_{yz}yz).$$

In turn, we have (taking conditionals and dropping constants)

$$\begin{aligned}
P(X=x,Y=y|Z=z) &\propto P(X=x,Y=y,Z=z) \\
&\propto \exp(\alpha_x x + \alpha_y y + \alpha_z z + \alpha_{xy}xy + \alpha_{xz}xz + \alpha_{yz}yz) \\
&\propto \exp(\alpha_x x + \alpha_y y + \alpha_{xy}xy + \alpha_{xz}xz + \alpha_{yz}yz) \cdot C(z).
\end{aligned}$$

In turn, we have

$$\begin{aligned}
\theta_{XY|Z} &= \frac{P(X=0,Y=0|Z=z)\cdot P(X=1,Y=1|Z=z)}{P(X=1,Y=0|Z=z)\cdot P(X=0,Y=1|Z=z)} \\
&= \frac{(\exp(0)C(z))\cdot((\exp(\alpha_x+\alpha_y+\alpha_{xy}+\alpha_{xz}z+\alpha_{yz}z)\cdot C(z))}{(\exp(\alpha_x+\alpha_{xz}z)\cdot C(z))\cdot(\exp(\alpha_y+\alpha_{yz}z)\cdot C(z))} \\
&= \frac{\exp(\alpha_x+\alpha_y+\alpha_{xy}+\alpha_{xz}z+\alpha_{yz}z)}{\exp(\alpha_x+\alpha_{xz}z+\alpha_y+\alpha_{yz}z)} \\
&= \exp(\alpha_{xy}).
\end{aligned}$$

This does not depend on z. The similar result holds, without loss of generality and by identical procedure, for the $\theta_{XZ|Y}$ and $\theta_{ZY|X}$ cases as well.

5.2

First, solve for $P(X=x|Z=z)$ by integrating out Y, i.e.

$$\begin{aligned}
P(X=x,Z=z) &\propto \sum_y \exp(\alpha_x x + \alpha_y y + \alpha_z z + \underbrace{\alpha_{xy}xy}_{0} + \alpha_{xz}xz + \alpha_{yz}yz) \\
&\propto \exp(\alpha_x x + \alpha_z z + \alpha_{xz}xz)\sum_y \exp(\alpha_y y + \alpha_{xy}xy + \alpha_{yz}yz) \\
&\propto \exp(\alpha_x x + \alpha_z z + \alpha_{xz}xz) g(y,z),
\end{aligned}$$

where g is the function of y,z necessary to complete the marginalization. Critically, ote that g has no x dependence, which allows it to drop off like a constant (w.r.t. x in the posterior calculation below). Then, we have

$$\begin{aligned}
P(X=x|Z=z) &\propto P(X=x,Z=z) \\
&\propto \exp(\alpha_x x + \alpha_z z + \alpha_{xz}xz) g(y,z) \\
&\propto \exp(\alpha_x x + \alpha_{xz}xz).
\end{aligned}$$

Subsequently, and more straightforwardly, we have for $P(X|Y,Z)$ (drop Y's and Z's as constants)

$$\begin{aligned}
P(X=x|Y=y,Z=z) &\propto P(X=x,Y=y,Z=z) \\
&\propto \exp(\alpha_x x + \alpha_y y + \alpha_z z + \alpha_{xy}xy + \alpha_{xz}xz + \alpha_{yz}yz) \\
&\propto \exp(\alpha_x x + \underbrace{\alpha_{xy}}_{0} xy + \alpha_{xz}xz) C(y,z) \\
&\propto \exp(\alpha_x x + \alpha_{xz}xz),
\end{aligned}$$

giving $P(X|Y,Z) = P(X|Z)$ and conditional independence, as desired.

Lastly, for the joint distribution $X,Y|Z=0$ and $X,Y|Z=1$ with the X,Y interaction zero'd (that's how I'm interpreting the question, it's a touch open ended). Using the results above, we have

$$\begin{aligned}
P(X=x,Y=y|Z=z) &\propto P(X=x,Y=y,Z=z) \\
&\propto \exp(\alpha_x x + \alpha_y y + \alpha_z z + \underbrace{\alpha_{xy}}_{0} xy + \alpha_{xz} xz + \alpha_{yz} yz) \\
&\propto \exp(\alpha_x x + \alpha_y y + \alpha_{xz} xz + \alpha_{yz} yz).
\end{aligned}$$

When $z=1$, this is

$$P(X=x,Y=y|Z=z) \propto \exp(\alpha_x x + \alpha_y y + \alpha_{xz} x + \alpha_{yz} y),$$

and when $z=0$, this is

$$P(X=x,Y=y|Z=z) \propto \exp(\alpha_x x + \alpha_y y).$$

5.3

Now, we generalize to V binary random variables in the joint. For fluidity of notation but without loss of generality, let X,Y remain the variables we wish to keep on the LHS of the conditional, and $Z_1, Z_2, \ldots, Z_{V-2}$ the "nuisance" parameters for the RHS of the conditional. Just as before, we will have

$$\begin{aligned}
P(X=x,Y=y|Z_1=z_1,\ldots,Z_{V-2}=z_{V-2}) &\propto P(X=x,Y=y,Z_1=z_1,\ldots,Z_{V-2}=z_{V-2}) \\
&\propto \exp\left(\alpha_x x + \alpha_y y + \alpha_{xy} xy + \sum_{j=1}^{V-2} \alpha_{z_j} z_j + \sum_{j=1}^{V-2} \alpha_{xz_j} xz_j + \sum_{j=1}^{V-2} \alpha_{yz_j} yz_j\right) \\
&\propto \exp\left(\alpha_x x + \alpha_y y + \alpha_{xy} xy + \sum_{j=1}^{V-2} \alpha_{xz_j} xz_j + \sum_{j=1}^{V-2} \alpha_{yz_j} yz_j\right) \cdot C(z_1,\ldots,z_{V-2}).
\end{aligned}$$

And as before

$$\begin{aligned}
\theta_{XY|Z_1,\ldots,Z_{V-2}} &= \frac{P(X=0,Y=0|Z_1,\ldots,Z_{V-2}) \cdot P(X=1,Y=1|Z_1,\ldots,Z_{V-2})}{P(X=1,Y=0|Z_1,\ldots,Z_{V-2}) \cdot P(X=0,Y=1|Z_1,\ldots,Z_{V-2})} \\
&= \frac{1 \cdot C(z) \cdot \exp\left(\alpha_x x + \alpha_y y + \alpha_{xy} + \sum_{j=1}^{V-2} \alpha_{xz_j} z_j + \sum_{j=1}^{V-2} \alpha_{yz_j} z_j\right) C(z)}{\exp\left(\alpha_x x + \sum_{j=1}^{V-2} \alpha_{xz_j} z_j\right) C(z) \cdot \exp\left(\alpha_y y + \sum_{j=1}^{V-2} \alpha_{yz_j} z_j\right) C(z)} \\
&= \exp(\alpha_{xy} xy),
\end{aligned}$$

as before. A result following identical mechanics holds for conditional independence (assuming $\alpha_{xy} = 0$ again), specifically

$$\begin{aligned}
P(X=x, Z_1=z_1,\ldots Z_{V-2}=z_{v-2}) &\propto \exp\left(\alpha_x x + \underbrace{\alpha_{xy}}_{0} xy + \sum_{j=1}^{V-2} \alpha_{z_j} z_j + \sum_{j=1}^{V-2} \alpha_{xz_j} xz_j\right) \sum_y \exp\left(\alpha_y y + \sum_{j=1}^{V-2} \alpha_{yz_j} yz_j\right) \\
&\propto \exp\left(\alpha_x x + \sum_{j=1}^{V-2} \alpha_{z_j} z_j + \sum_{j=1}^{V-2} \alpha_{xz_j} xz_j\right) g(z_1,\ldots,g_{V-2},y).
\end{aligned}$$

Where g describes the integrating function to margin out y. Again, note that x has no dependence on $g(z, y)$. In turn, it follows that

$$\begin{aligned}
P(X|Z_1, \dots Z_{V-2}) &\propto P(X, Z_1, \dots Z_{V-2}) \\
&\propto \exp\left(\alpha_x x + \sum_{j=1}^{V-2} \alpha_{z_j} z_j + \sum_{j=1}^{V-2} \alpha_{xz_j} x z_j\right) g(z_1, \dots, g_{V-2}, y) \\
&\propto \exp\left(\alpha_x x + \sum_{j=1}^{V-2} \alpha_{xz_j} x z_j\right).
\end{aligned}$$

Then, for $P(X|Y, Z_1, \dots, Z_{V-2})$, we again just drop constants w.r.t. x:

$$\begin{aligned}
P(X|Y, Z_1, \dots, Z_{V-2}) &\propto P(X, Y, Z_1, \dots, Z_{V-2}) \\
&\propto \exp\left(\alpha_x x + \alpha_y y + \underbrace{\alpha_{xy}}_{0} xy + \sum_{j=1}^{V-2} \alpha_{z_j} z_j + \sum_{j=1}^{V-2} \alpha_{xz_j} x z_j + \sum_{j=1}^{V-2} \alpha_{yz_j} y z_j\right) \\
&\propto \exp\left(\alpha_x x + \sum_{j=1}^{V-2} \alpha_{xz_j} x z_j + \right),
\end{aligned}$$

showing that $P(X|Z_1, \dots Z_{V-2}) = P(X|Y, Z_1, \dots Z_{V-2})$, as desired.

5.4

First, suppose that each of X, Y,and Z fall into, respectively, k_x, k_y and k_z categories. We encode

- $\vec{X} \in \mathbb{Z}_2^{k_x}$
- $\vec{Y} \in \mathbb{Z}_2^{k_y}$
- $\vec{Z} \in \mathbb{Z}_2^{k_z}$,

i.e. a one-hot-encoding for each of the categorical responses. Then, define

- $\vec{\alpha}_x \in \mathbb{R} k_x$
- $\vec{\alpha}_y \in \mathbb{R} k_y$
- $\vec{\alpha}_z \in \mathbb{R} k_z$,

these are the non-interacted coefficients. Lastly, define tuples

- $\alpha_{xy} = (\vec{\alpha}_{xy;L}, \vec{\alpha}_{xy;R}) \in \mathbb{R}^{k_x \times k_y}$
- $\alpha_{xz} = (\vec{\alpha}_{xz;L}, \vec{\alpha}_{xz;R}) \in \mathbb{R}^{k_x \times k_z}$
- $\alpha_{yz} = (\vec{\alpha}_{yz;L}, \vec{\alpha}_{yz;R}) \in \mathbb{R}^{k_y \times k_z}$.

This prepares us to interact via outer products, i.e.

$$\begin{aligned}
P(\vec{X} = \vec{x}, \vec{Y} = \vec{y}, \vec{Z} = \vec{z}) &\propto \exp\left(\vec{\alpha}_x^T \vec{x} + \vec{\alpha}_y^T \vec{y} + \vec{\alpha}_z^T \vec{z} + \alpha_{xy}(\vec{x}, \vec{y}) + \alpha_{xz}(\vec{x}, \vec{z}) + \alpha_{zy}(\vec{y}, \vec{z})\right) \\
&\propto \exp\left(\vec{\alpha}_x^T \vec{x} + \vec{\alpha}_y^T \vec{y} + \vec{\alpha}_z^T \vec{z} + \vec{\alpha}_{xy;L}^T \vec{x} \vec{y}^T \alpha_{xy;R} + \vec{\alpha}_{xz;L}^T \vec{x} \vec{z}^T \alpha_{xz;R} + \vec{\alpha}_{yz;L}^T \vec{y} \vec{z}^T \alpha_{yz;R}.\right).
\end{aligned}$$

By following the algebraic mechanisms above in 1/2 – albeit at the vector/tuple levels of parameterization now – we can recover the same results. Moreover, the one-hot encoding effectively reduces this problem to that posed in 3, with a massive number of binarized variables.

To better understand this setup, consider the original binary setup under this formulation. If we construct (borrowing from our old single-variable/binarized notation)

- $X = 0 \iff \vec{X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $X = 1 \iff \vec{X} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- $Y = 0 \iff \vec{Y} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $Y = 1 \iff \vec{Y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- $Z = 0 \iff \vec{Z} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $Z = 1 \iff \vec{Z} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- $\vec{\alpha}_x = \begin{bmatrix} 0 \\ \alpha_x \end{bmatrix}, \vec{\alpha}_y = \begin{bmatrix} 0 \\ \alpha_y \end{bmatrix}, \vec{\alpha}_z = \begin{bmatrix} 0 \\ \alpha_z \end{bmatrix}$
- $\vec{\alpha}_{xy;L} = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $\vec{\alpha}_{xy;R} = \begin{bmatrix} 0 & \alpha_{xy} \end{bmatrix}$
- $\vec{\alpha}_{xz;L} = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $\vec{\alpha}_{xz;R} = \begin{bmatrix} 0 & \alpha_{xz} \end{bmatrix}$
- $\vec{\alpha}_{yz;L} = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $\vec{\alpha}_{yz;R} = \begin{bmatrix} 0 & \alpha_{yz} \end{bmatrix}$,

then the joint PMF will have the form

$$\begin{aligned} P(\vec{X} = \vec{x}, \vec{Y} = \vec{y}, \vec{Z} = \vec{z}) &\propto \exp\Big(\vec{\alpha}_x^T \vec{x} + \vec{\alpha}_y^T \vec{y} + \vec{\alpha}_z^T \vec{z} + \vec{\alpha}_{xy;L}^T \vec{x}\vec{y}^T \alpha_{xy;R} + \vec{\alpha}_{xz;L}^T \vec{x}\vec{z}^T \alpha_{xz;R} + \vec{\alpha}_{yz;L}^T \vec{y}\vec{z}^T \alpha_{yz;R}.\Big) \\ &= \exp(\alpha_x x_1 + \alpha_y y_1 + \alpha_z z_1 + \alpha_{xy} x_1 y_1 + \alpha_{xz} x_1 z_1 + \alpha_{yz} y_1 z_1.), \end{aligned}$$

which is our original PMF. Note that the interactions follow from expansions/outer products of the form

$$\begin{aligned} \vec{\alpha}_{xy;L}^T \vec{x}\vec{y}^T \vec{\alpha}_{xy;R} &= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \begin{bmatrix} y_0 & y_1 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha_{xy} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 y_0 & x_0 y_1 \\ x_1 y_0 & x_1 y_1 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha_{xy} \end{bmatrix} \\ &= \begin{bmatrix} x_1 y_0 & x_1 y_1 \end{bmatrix} \begin{bmatrix} 0 \\ \alpha_{xy} \end{bmatrix} \\ &= \alpha_{xy} x_1 y_1. \end{aligned}$$

5.5

Now, suppose we have

$$X \sim MVN(0_k, \Sigma),$$

with

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k. \end{bmatrix}$$

Without loss of generality, suppose we are interested in x_1 and secondarily x_2, with the remainder of the x_i on the right side of the conditional. For the joint density of this vector, we have

$$\begin{aligned} p(x_1, \dots, x_k) &\propto \exp\Big(-\frac{1}{2}\vec{x}^T \Sigma^{-1} \vec{x}\Big) \\ &\propto \exp\Big(-\frac{1}{2}\sum_{i=1}^k \sum_{j=1}^k x_i \theta_{ij} x_j\Big) \\ &\propto \exp\Big(-\frac{1}{2}\sum_{i=1}^k x_i^2 \theta_{ii} - \sum_{i \neq j}^k x_i x_j \theta_{ij}.\Big) \end{aligned}$$

Now, for the full conditional, we have

$$\begin{aligned}
p(x_1|x_2,\dots,x_k) &\propto \exp\Big(-\frac{1}{2}\sum_{i=1}^{k}x_i^2\theta_{ii} - \sum_{i\neq j}^{k}x_ix_j\theta_{ij}.\Big)\\
&\propto \exp\Big(-\frac{1}{2}x_1^2\theta_{11} - \sum_{j\geq 2}^{k}x_1x_j\theta_{1j}.\Big)\\
&\propto \exp\Big(-\frac{1}{2}x_1^2\theta_{11} - \underbrace{x_1x_2\theta_{12}}_{0} - \sum_{j\geq 3}^{k}x_1x_j\theta_{1j}.\Big)\\
&\propto \exp\Big(-\frac{1}{2}x_1^2\theta_{11} - \sum_{j\geq 3}^{k}x_1x_j\theta_{1j}.\Big).
\end{aligned}$$

Then, we have (integrating out the x_2 component),

$$p(x_1,x_3,x_4,\dots,x_k) \sim MVN(\vec{0}_{k-1},\Sigma'),$$

where Σ' is just the Σ matrix with the second (for x_2) row and column deleted. This property follows from the marginal distribution property of the MVN, in which you can drop dimensions and the other means covariances remain (`https://en.wikipedia.org/wiki/Multivariate_normal_distribution#Marginal_distributions`). Accordingly, we have

$$\begin{aligned}
p(x_1|x_3,x_4,\dots,x_k) &\propto p(x_1,x_3,x_4,\dots,x_k)\\
&\propto \exp\Big(-\frac{1}{2}\sum_{i=1,3,\dots,k}x_i^2\theta_{ii} - \sum_{i\neq j,i\neq 2,j\neq 2}^{k}x_ix_j\theta_{ij}.\Big)\\
&\propto \exp\Big(-\frac{1}{2}x_1^2\theta_{11} - \sum_{j\geq 3}^{k}x_1x_j\theta_{1j}.\Big),
\end{aligned}$$

as desired. We have shown conditional independence, i.e.

$$p(x_1|x_3,x_4,\dots,x_k) = p(x_1|x_2,x_3,x_4,\dots,x_k)$$

for this covariance structure.

Notably, there is an obvious similarity between the conditional independence structure here and that of the categorical RV described in parts I-IV. Specifically, given (without loss) an arbitrary choice of X, Y parameters, we have that when the interaction parameter between two of X and Y (in the categorical case, $\alpha_{xy} = 0$; in the case of the MVN, θ_{xy}) is zeroed out, conditional independence holds, that is, X and Y are conditionally independent given Z or the rest of the parameters $Z_1, Z_2, \dots$.