QUESTION 1

(i) This is false.

The joint entropy is $S(X \cap Y) = -\sum_{x,y} p(x \cap y) \ln p(x \cap y)$. When the events are independent the probabilities factor as a product of marginals, $p(x \cap y) = p(x)p(y)$. It follows that the entropy is additive, $S(X \cap Y) = S(X) + S(Y)$.

(ii) This is false.

The conditional entropy is defined as $S(X|Y) = \sum_y p(y)S(X|y)$, where $S(X|y) = -\sum_x p(x|y) \ln p(x|y)$. One may state that the conditional probabilities $p(x|x)$ are evidently unity. Otherwise, one might say that in general $p(x|y) = p(x \cap y)/p(y)$ and the joint probabilities $p(x \cap x)$ are evidently equal to $p(x)$. More generally $p(x|x')$ is 1 if $x = x'$ and zero otherwise. It follows that $S(X|X) = 0$; there is no uncertainty in X if you already know X.

(iii) This is true.

The mutual information is defined as $I(X, Y) = S(X) + S(Y) - S(X \cap Y)$. On the basis that $p(x \cap x')$ equals $p(x)$ if $x = x'$ and zero otherwise, one finds $S(X \cap X) = S(X)$ and hence that $I(X, Y) = S(X)$. Knoweldge of X tells you everything about X.

(iv) This is false.

Let f be a constant function. Then $S\big(f(X)\big) = 0$ independently of $S(X)$.

QUESTION 2

We will refer to the following table of possible outcomes and values of the properties P and Q.

outcome	1	2	3	4	5	6
P	0	1	1	0	1	0
Q	0	1	0	1	0	1

(i) From the table we see that the probability of the outcome being prime and the probability of it being non-prime are both $\frac{1}{2}$. It follows that the Shannon entropy is

$$S(P) = -\sum_{\alpha} p_\alpha \ln(p_\alpha) = -\left[\frac{1}{2}\ln\left(\frac{1}{2}\right) + \frac{1}{2}\ln\left(\frac{1}{2}\right)\right] = \ln(2).$$

Similarly, the probability of the outcome being even and the probability of it being odd are both $\frac{1}{2}$ and it follows that $S(Q) = \ln(2)$ as well.

(ii) There are four joint events; $\{P, Q\} = \{0, 0\}, \{0, 1\}, \{1, 0\}, \{1, 1\}$. Their probabilities are

$$p_{\{0,0\}} = \frac{1}{6}, \quad p_{\{0,1\}} = \frac{1}{3}, \quad p_{\{1,0\}} = \frac{1}{3}, \quad p_{\{1,1\}} = \frac{1}{6},$$

and it follows immediately that the joint entropy is

$$S(P, Q) = \frac{1}{6}\ln(6) + \frac{1}{3}\ln(3) + \frac{1}{3}\ln(3) + \frac{1}{6}\ln(6) = \ln(3) + \frac{1}{3}\ln(2).$$

(iii) The conditional probabilities are given by $p(P|Q) = p(P \cap Q)/p(Q)$ and $p(Q|P) = p(P \cap Q)/p(P)$. We quickly find that

$$p(P=0|Q=0) = \frac{1}{3}, \quad p(P=1|Q=0) = \frac{2}{3}\,; p(P=0|Q=1) = \frac{2}{3}, \quad p(P=1|Q=1) = \frac{1}{3};$$
$$p(Q=0|P=0) = \frac{1}{3}, \quad p(Q=1|P=0) = \frac{2}{3}\,; p(Q=0|P=1) = \frac{2}{3}, \quad p(Q=1|P=1) = \frac{1}{3}.$$

We can then use these conditional probabilities in the expression for the Shannon entropy and find

$$S(P|Q=0) = -\sum p(P|Q=0)\ \ln\big(p(P|Q=0)\big) = \frac{1}{3}\ln(3) + \frac{2}{3}\ln\left(\frac{3}{2}\right) = \ln(3) - \frac{2}{3}\ln(2).$$

The same value is found for $S(P|Q=1)$, $S(Q|P=0)$ and $S(Q|P=1)$. Hence the conditional entropies are

$$S(P|Q) = \sum_q p(P|Q=q) S(P|Q=q) = \ln(3) - \frac{2}{3}\ln(2),$$

and the same for $S(Q|P)$.

We can readily verify that the relation $S(P|Q) = S(P, Q) - S(Q)$ is indeed satisfied

$$S(P|Q) = \ln(3) - \frac{2}{3}\ln(2) = \ln(3) + \frac{1}{3}\ln(2) - \ln(2) = S(P, Q) - S(Q).$$

Likewise we can verify the relation $S(Q|P) = S(P, Q) - S(P)$ from an identical check.

(iv) The mutual information is given by

$$I(P,Q) = S(P) + S(Q) - S(P,Q) = 2\,\ln(2) - \ln(3) - \frac{1}{3}\,\ln(2) = \frac{5}{3}\,\ln(2) - \ln(3).$$

We can check the expected relation $I(P,Q) = S(P) - S(P|Q)$ as follows

$$I(P,Q) = \frac{5}{3}\,\ln(2) - \ln(3) = \ln(2) - \left(\ln(3) - \frac{2}{3}\,\ln(2)\right) = S(P) - S(P|Q).$$

A similar check verifies the relation $I(P,Q) = S(Q) - S(Q|P)$.

We are told that an unfair die on average rolls a 4, or $\sum_{\alpha=1}^{6} \alpha p_\alpha = 4$. We can use maximum entropy inference to estimate the probabilities p_α subject to this constraint on the average. Using Lagrange multipliers λ_0 and λ_1 for the constraint on the total probability being 1 and the average roll being 4 we have

$$\begin{aligned} \mathrm{d}\Big[-\sum_\alpha p_\alpha\,\ln(p_\alpha) - \lambda_0\Big(\sum_\alpha p_\alpha - 1\Big) - \lambda_1\Big(\sum_\alpha \alpha p_\alpha - 4\Big)\Big] &= 0, \\ \Rightarrow \qquad -\sum_\alpha \Big[\ln(p_\alpha) + 1 + \lambda_0 + \lambda_1\alpha\Big]\,\mathrm{d}p_\alpha &= 0, \end{aligned}$$

and it follows that the probabilities are given by

$$p_\alpha = \frac{1}{Z}\,\mathrm{e}^{-\lambda_1\alpha}, \qquad Z = \sum_\alpha \mathrm{e}^{-\lambda_1\alpha}.$$

The Lagrange multiplier λ_1 can be determined from the constraint on the average roll, giving

$$\begin{aligned} \sum_\alpha (\alpha - 4)p_\alpha = 0 &= 2\,\mathrm{e}^{-6\lambda_1} + \mathrm{e}^{-5\lambda_1} - \mathrm{e}^{-3\lambda_1} - 2\,\mathrm{e}^{-2\lambda_1} - 3\,\mathrm{e}^{-\lambda_1}, \\ \text{or} \qquad 2x^5 + x^4 - x^2 - 2x - 3 &= 0, \quad x = \mathrm{e}^{-\lambda_1}. \end{aligned}$$

The solution may be found numerically. For instance, I used the FindRoot function in Mathematica, which gave $x = 1.1908$. I then obtain for the probabilities

$$p_1 = 0.10, \quad p_2 = 0.12, \quad p_3 = 0.15, \quad p_4 = 0.17, \quad p_5 = 0.21, \quad p_6 = 0.25,$$

so that the maximum entropy estimate for the probability of getting a 6 is $p_6 = 0.25$.

QUESTION 3

(i) There are two main ways to do this; using transfer matrices or by defining a convenient expansion of the Boltzmann factor as we did for the Ising model in the lectures. For the first approach, the transfer matrix is a $q\times q$ matrix with entries $\mathrm{e}^{\beta J}$ on the diagonal and 1 elsewhere. One can check that the vector with all entries 1 is an eigenvector with eigenvalue $\mathrm{e}^{\beta J}+(q-1)$ and that the vector with entries $1,-1,0,\ldots,0$ is an eigenvector with eigenvalue $\mathrm{e}^{\beta J}-1$. In fact the symmetry of the problem tells us that all of the other eigenvectors have this same eigenvalue. Hence the partition function is

$$Z=\operatorname{tr}T^N=\sum_a \lambda_a^N=\Big(\mathrm{e}^{\beta J}+q-1\Big)^N+(q-1)\Big(\mathrm{e}^{\beta J}-1\Big)^N.$$

For the second approach we write

$$Z=\sum_{\{s_i\}}\mathrm{e}^{\beta J\sum_i\delta_{s_i,s_{i+1}}}=\sum_{\{s_i\}}\prod_i C\Big[1+x\big(q\delta_{s_i,s_{i+1}}-1\big)\Big],$$

where the second expression defines both C and x. Taking separately $s_{i+1}=s_i$ and $s_{i+1}\neq s_i$ we find

$$\begin{aligned} &\mathrm{e}^{\beta J}=C\big[1+x(q-1)\big], && 1=C\big[1-x\big],\\ \Longrightarrow\quad & x=\frac{\mathrm{e}^{\beta J}-1}{\mathrm{e}^{\beta J}+q-1}, && C=\frac{\mathrm{e}^{\beta J}+q-1}{q}.\end{aligned}$$

When we multiply out the product we will encounter sums over the possible values of the spins of functions of the form $q\delta_{s_i,s_{i+1}}-1$ and products of such. These are given by

$$\sum_{s_i=1}^{q}\big(q\delta_{s_i,s_j}-1\big)=0,$$

$$\sum_{s_j=1}^{q}\big(q\delta_{s_i,s_j}-1\big)\big(q\delta_{s_j,s_k}-1\big)=q\big(q\delta_{s_i,s_k}-1\big),$$

$$\sum_{s=1}^{q}\big(q\delta_{s,s}-1\big)=q(q-1).$$

With these we can compute the partition function and find

$$\begin{aligned} Z&=C^Nq^N+C^Nx^Nq^{N-1}q(q-1)=(Cq)^N+(Cq)^Nx^N(q-1),\\ &=\Big(\mathrm{e}^{\beta J}+q-1\Big)^N+(q-1)\Big(\mathrm{e}^{\beta J}-1\Big)^N.\end{aligned}$$

(ii) The partition function is $Z=\mathrm{e}^{-\beta F}$, so the free energy per site is

$$\begin{aligned}\frac{F}{N}&=-k_{\mathrm{B}}T\,\ln\Big(\mathrm{e}^{\beta J}+q-1\Big)+\frac{1}{N}\ln\left[1+(q-1)\left(\frac{\mathrm{e}^{\beta J}-1}{\mathrm{e}^{\beta J}+q-1}\right)^N\right],\\ &\to -k_{\mathrm{B}}T\,\ln\Big(\mathrm{e}^{\beta J}+q-1\Big).\end{aligned}$$

(iii) To find the correlation length we may compute the expectation value of $q\delta_{s_k,s_{k+r}} - 1$, which may be done either with transfer matrices, or by the expansion technique used for the partition function. Pursuing the latter approach we find

$$
\begin{aligned}
\left\langle q\delta_{s_k,s_{k+r}} - 1 \right\rangle &= \frac{1}{Z}\sum_{\{s_i\}}\prod_i \left(q\delta_{s_k,s_{k+r}} - 1\right) C\Big[1 + x\left(q\delta_{s_i,s_{i+1}} - 1\right)\Big], \\
&= \frac{1}{Z}\Big[C^N q^r.x^{N-r}q^{N-r-1}q(q-1) + C^N q^{N-r}.x^r q^{r-1}q(q-1)\Big], \\
&= \frac{(q-1)x^{N-r} + (q-1)x^r}{1+(q-1)x^N}, \\
&\to (q-1)x^r, \\
&= \mathrm{e}^{r\ln x + \ln(q-1)},
\end{aligned}
$$

which gives a correlation length of $(\ln 1/x)^{-1}$ or

$$
\xi = \frac{1}{\ln \frac{\mathrm{e}^{\beta J}+q-1}{\mathrm{e}^{\beta J}-1}}.
$$

QUESTION 4

(i) We write the argument of the logarithm as

$$\begin{aligned}(1+x^2)^2 - 2x(1-x^2)\big[\cos(q_1)+\cos(q_2)\big] &= 1+2x^2+x^4-4x(1-x^2)\\ &\quad + 4x(1-x^2)\big[\sin^2(q_1/2)+\sin^2(q_2/2)\big],\\ &= \big(x^2+2x-1\big)^2 + 4x(1-x^2)\big[\sin^2(q_1/2)+\sin^2(q_2/2)\big],\end{aligned}$$

and both terms are non-negative as x takes values in $[0,1]$.

(ii) The argument of the logarithm vanishes when $q_1 = q_2 = 0$ and $x^2+2x-1=0$. This latter is solved by

$$x = x_c = \sqrt{2}-1.$$

In the lectures we found using Kramers-Wannier duality that $\sinh(2\beta_c J) = 1$. Using standard hyperbolic trig identities this can be written as

$$1 = 2\sinh(\beta_c J)\cosh(\beta_c J) = 2\tanh(\beta_c J)\cosh^2(\beta_c J),$$

so that it is equivalent to

$$2\tanh(\beta_c J) = \mathrm{sech}^2(\beta_c J) = 1-\tanh^2(\beta_c J),$$

or $2x_c = 1-x_c^2$. Evidently the two expressions are the same.

(iii) Write $x = x_c + (x-x_c)$ and expand assuming $x-x_c$ is small. We find

$$\big(x^2+2x-1\big)^2+4x(1-x^2)\big[\sin^2(q_1/2)+\sin^2(q_2/2)\big] = \big(2\sqrt{2}(x-x_c)\big)^2+x_c(1-x_c^2)\big[q_1^2+q_2^2\big]+\text{h.o.t.},$$

where h.o.t. means 'higher order terms'. Hence the integral in the expression for the partition function can be written

$$\begin{aligned}I &= \int_{-\pi}^{\pi}\frac{dq_1}{2\pi}\int_{-\pi}^{\pi}\frac{dq_2}{2\pi}\,\ln\Big[(1+x^2)^2-2x(1-x^2)\big(\cos(q_1)+\cos(q_2)\big)\Big],\\ &= \int\frac{d^2q}{(2\pi)^2}\,\ln\Big[8(x-x_c)^2+2x_c^2\,q^2\Big]+\text{regular},\end{aligned}$$

where 'regular' means that the integral is finite, even setting $x = x_c$. We note that the divergent contribution to the integral comes from close to the origin, $q \approx 0$, and so split the domain of integration into a circle, of radius π, say, about the origin, and everything else. The latter contribution, being at non-zero q is regular, so we can absorb it into the terms that we are neglecting. Hence, the leading contribution is

$$\begin{aligned}I &= \int_0^{\pi}\frac{q\,dq}{2\pi}\,\ln\Big[8(x-x_c)^2+2x_c^2\,q^2\Big]+\text{regular},\\ &= \frac{1}{2\pi}\frac{1}{4x_c^2}\Big(8(x-x_c)^2+2x_c^2\,q^2\Big)\ln\Big[8(x-x_c)^2+2x_c^2\,q^2\Big]\Bigg|_{q=0}^{\pi} - \int_0^{\pi}\frac{dq}{2\pi}\,q+\text{regular},\\ &= \frac{(x-x_c)^2}{\pi x_c^2}\ln\big[8(x-x_c)^2\big]+\text{regular},\end{aligned}$$

where in the second line we have integrated by parts (cleverly) and in the third continue to place unimportant contributions into the 'regular' part. We see that the leading singular behaviour of the free energy at the transition comes from the contribution

$$F \sim \frac{(x-x_c)^2}{\pi x_c^2}\ln|x-x_c|.$$

Since the heat capacity is proportional to the second derivative of F with respect to T, and $x-x_c$ and $T-T_c$ are related in a generic fashion (use the implicit function theorem), we find that the heat capacity exhibits a logarithmic singularity at the transition, $C_V \sim \ln|T-T_c|$.