Homework 6

Due 3/7

---

1. (a) The joint entropy is (where $\log = \log_2$)

$$
\begin{aligned}
H(V,T) &= \sum_{V,T} P(V,T)\log\left(\frac{1}{P(V,T)}\right) \\
&= \left[\frac{6}{16}\log(16) + \frac{4}{32}\log(32) + \frac{2}{8}\log(8) + \frac{1}{4}\log(4)\right] = \frac{54}{16} \\
H(V,T) &= \boxed{3.38 \text{ bits}}
\end{aligned}
$$

Given the marginal probability

$$
\begin{aligned}
P(V = \text{Sunny}) &= \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{4}{16} = \frac{1}{4} \\
P(V = \text{Cloudy \& dry}) &= \frac{1}{16} + \frac{1}{8} + \frac{1}{32} + \frac{1}{32} = \frac{8}{32} = \frac{1}{4} \\
P(V = \text{Cloudy \& rain}) &= \frac{1}{4} \\
P(V = \text{Cloudy \& snow}) &= \frac{1}{4}
\end{aligned}
$$

marginal entropy of V is

$$
\begin{aligned}
H(V) &= \sum_{V} P(V)\log\left(\frac{1}{P(V)}\right) \\
&= \frac{1}{4}\log(4) + \frac{1}{4}\log(4) + \frac{1}{4}\log(4) + \frac{1}{4}\log(4) \\
H(V) &= \boxed{2 \text{ bits}}
\end{aligned}
$$

And given the marginal probability

$$
\begin{aligned}
P(T = \text{Miserably Cold}) &= \frac{1}{16} + \frac{1}{16} + \frac{1}{8} + \frac{1}{4} = \frac{1}{2} \\
P(T = \text{Very Cold}) &= \frac{1}{4} \\
P(T = \text{Cold}) &= \frac{1}{8} \\
P(T = \text{Chilly}) &= \frac{1}{8}
\end{aligned}
$$

marginal entropy of T is

$$
\begin{aligned}
H(T) &= \sum_{T} P(T)\log\left(\frac{1}{P(T)}\right) \\
&= \frac{1}{2}\log(2) + \frac{1}{4}\log(4) + \frac{1}{8}\log(8) + \frac{1}{8}\log(8) \\
H(T) &= \boxed{1.75 \text{ bits}}
\end{aligned}
$$

(b) The conditional entropy of T given $V = v$ is

$$
H(T|V=v) = \sum_{T} P(T|V=v)\log\left(\frac{1}{P(T|V=v)}\right)
$$

and from Bayes' theorem

$$P(T|V=v) = \frac{P(V=v,T)}{P(V=v)}$$

So for $v = \text{Sunny}$:

$$H(T|V=\text{Sunny}) = \frac{1}{4}\log(4) + \frac{1}{4}\log(4) + \frac{1}{4}\log(4) + \frac{1}{4}\log(4) = \boxed{2 \text{ bits}}$$

For $v = \text{Cloudy \& dry}$:

$$H(T|V=\text{Cloudy \& dry}) = \frac{1}{4}\log(4) + \frac{1}{2}\log(2) + \frac{1}{8}\log(8) + \frac{1}{8}\log(8) = \boxed{1.75 \text{ bits}}$$

For $v = \text{Cloudy \& rain}$:

$$H(T|V=\text{Cloudy \& rain}) = \frac{1}{2}\log(2) + \frac{1}{4}\log(4) + \frac{1}{8}\log(8) + \frac{1}{8}\log(8) = \boxed{1.75 \text{ bits}}$$

For $v = \text{Cloudy \& snow}$:

$$H(T|V=\text{Cloudy \& snow}) = \log(1) = \boxed{0 \text{ bits}}$$

this makes sense since its *always* miserably cold given it is Cloudy & snowing.
(c) The conditional entropy as an average

$$\begin{aligned}
H(T|V) &= \sum_V P(V)[H(T|V=v)] \\
&= \frac{1}{4}H(T|V=v) \\
H(T|V) &= \frac{1}{4}(2 + 1.75 + 1.75 + 0) = \boxed{1.38 \text{ bits}}
\end{aligned}$$

(d) Using product rule on the joint entropy:

$$\begin{aligned}
H(V,T) &= \sum_{V,T} P(V,T)\log\left(\frac{1}{P(T|V)P(T)}\right) \\
&= \sum_{V,T} P(V,T)\log\left(\frac{1}{P(T|V)}\right) + \sum_{V,T} P(V,T)\log\left(\frac{1}{P(T)}\right)
\end{aligned}$$

and from sum the sum rule:

$$\begin{aligned}
P(T) &= \sum_V P(V,T) \\
&= \sum_V P(T|V)P(V)
\end{aligned}$$

so

$$H(V,T) = H(T|V) + H(T) \implies H(T|V) = H(V,T) - H(T) = 3.38 - 2 = 1.38 \text{ bits}$$

which confirms the result from part (c), and we can also see that

$$\begin{aligned}
&H(V,T) = H(T) + H(V|T) \\
\implies &H(V|T) = H(V,T) - H(T) = 3.38 - 1.75 = \boxed{1.63 \text{ bits}}
\end{aligned}$$

(e) The mutual information is

$$\begin{aligned}
I(V;T) &= H(V) - H(V|T) \quad \text{or} \quad H(T) - H(T|V) \\
&= 2 - 1.63 = \boxed{0.37 \text{ bits}}
\end{aligned}$$

2. (a) For a Gaussian defined by the PDF(Probability Density Function)

$$P(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

the entropy is (here log is the natural logarithm $\log_e = \ln$ i.e. unit of nats)

$$\begin{aligned}
H(P) &= -\int_{-\infty}^{\infty} P(x)\log(P(x))\,\mathrm{d}x \\
&= -\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \left(\frac{-x^2}{2\sigma^2} - \log\left(\sqrt{2\pi\sigma^2}\right)\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{8\pi\sigma^6}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x + \frac{\log\left(\sqrt{2\pi\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,\mathrm{d}x
\end{aligned}$$

and using some useful Gaussian integrals:

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-ax^2}\,\mathrm{d}x &= \sqrt{\frac{\pi}{a}} \\
\int_{-\infty}^{\infty} x^2 e^{-ax^2}\,\mathrm{d}x &= \frac{1}{2}\sqrt{\frac{\pi}{a^3}}
\end{aligned}$$

where $a = \frac{1}{2\sigma^2}$, so

$$\begin{aligned}
H(P) &= \frac{1}{\sqrt{8\pi\sigma^6}} \left[\frac{1}{2}\sqrt{8\pi\sigma^6}\right] + \frac{\log\left(\sqrt{2\pi\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \sqrt{2\pi\sigma^2} \\
&= \frac{1}{2} + \frac{1}{2}\log\left(2\pi\sigma^2\right) \\
&= \frac{1}{2}\left[1 + \log\left(2\pi\sigma^2\right)\right]
\end{aligned}$$

and $H(P)$ can be negative when

$$\begin{aligned}
& 1 + \log\left(2\pi\sigma^2\right) < 0 \\
\implies \sigma^2 < \frac{1}{2\pi e} & \quad \text{or} \quad \sigma < \frac{1}{\sqrt{2\pi e}}
\end{aligned}$$

(b) Since ξ and X are independent, the variance of $Y = \xi + X$ is the sum of the variances

$$\mathrm{Var}(Y) = \mathrm{Var}(\xi) + \mathrm{Var}(X) = \sigma_\xi^2 + \sigma_X^2$$

(c) From the Sum rule

$$P_Y(y) = \sum_{Z=z} P_\xi(\xi = z) P_X(X)$$

we can change the discrete case to a continuous one by integrating over all possible values of $\xi = z$ to find the probability density function $P_Y(y)$:

$$\begin{aligned}
P_Y(y) &= \int_{-\infty}^{\infty} P_\xi(\xi = z) P_X(X|\xi = z)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P_{X,\xi}(X, \xi = z)\,\mathrm{d}z
\end{aligned}$$

Since ξ and X are independent, $P_{X,\xi}(X,\xi) = P_X(X)P_\xi(\xi)$, and $X = Y - \xi$:

$$\begin{aligned}
P_Y(y) &= \int_{-\infty}^{\infty} P_X(X = y - z) P_\xi(\xi = z)\,\mathrm{d}z \\
&= \int_{-\infty}^{\infty} P_X(y - z) P_\xi(z)\,\mathrm{d}z
\end{aligned}$$

(d) We can just plug in the Gaussian PDFs for P_X and P_ξ where we assume the means are zero:

$$\begin{aligned}
P_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(y-z)^2}{2\sigma_X^2}} \frac{1}{\sqrt{2\pi\sigma_\xi^2}} e^{-\frac{z^2}{2\sigma_\xi^2}} \,\mathrm{d}z \\
&= \frac{1}{2\pi\sigma_X\sigma_\xi} \int_{-\infty}^{\infty} e^{-\frac{y^2-2yz+z^2}{2\sigma_X^2} - \frac{z^2}{2\sigma_\xi^2}} \,\mathrm{d}z \\
&= \frac{1}{2\pi\sigma_X\sigma_\xi} e^{-\frac{y^2}{2\sigma_X^2}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2\sigma_X^2} - \frac{z^2}{2\sigma_\xi^2} + \frac{yz}{\sigma_X^2}} \,\mathrm{d}z \\
&= \frac{1}{2\pi\sigma_X\sigma_\xi} \left(\frac{2\pi\sigma_X^2\sigma_\xi^2}{\sigma_X^2+\sigma_\xi^2}\right)^{1/2} e^{-\frac{y}{2(\sigma_X^2+\sigma_\xi^2)}} \\
&= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}} e^{-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)}}
\end{aligned}$$

which is also a Gaussian with our expected variance (add variances)!
(e) The mutual information is

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

so

$$\begin{aligned}
H(Y) &= -\int_{-\infty}^{\infty} P_Y(y)\log(P(y))\,\mathrm{d}y \\
&= -\frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)}} \log\left(\frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}} e^{-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)}}\right) \mathrm{d}y \\
&= -\frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)}} \left[-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)} - \log\left(\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}\right)\right] \mathrm{d}y \\
&= \frac{1}{\sqrt{8\pi(\sigma_X^2+\sigma_\xi^2)^3}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)}} \,\mathrm{d}y + \frac{\log\left(\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}\right)}{\sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2(\sigma_X^2+\sigma_\xi^2)}}
\end{aligned}$$

using

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-ay^2}\,\mathrm{d}x &= \sqrt{\frac{\pi}{a}} = \sqrt{2\pi(\sigma_X^2+\sigma_\xi^2)} \\
\int_{-\infty}^{\infty} y^2 e^{-ay^2}\,\mathrm{d}x &= \frac{1}{2}\sqrt{\frac{\pi}{a^3}} = \frac{1}{2}\sqrt{8\pi(\sigma_X^2+\sigma_\xi^2)^3}
\end{aligned}$$

where $a = \frac{1}{2(\sigma_X^2+\sigma_\xi^2)}$, so

$$H(Y) = \frac{1}{2} + \frac{1}{2}\log\left(2\pi(\sigma_X^2+\sigma_\xi^2)\right)$$

and

$$\begin{aligned}
-H(Y|X) &= -\int_{-\infty}^{\infty} P_X(x) H(Y|X=x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} P_X(x)\left[\int_{-\infty}^{\infty} P_Y(y|X=x)\log(P_Y(y|X=x))\,\mathrm{d}y\right]\mathrm{d}x
\end{aligned}$$

in the second integral we can use $y = x + \xi$ so

$$P_Y(y|X = x) = P_\xi(\xi = y - x|X = x)$$

and since ξ and X are independent

$$P_\xi(\xi = y - x|X = x) = P_\xi(\xi)$$

so

$$\begin{aligned} -H(Y|X) &= \int_{-\infty}^{\infty} P_X(x)\left[\int_{-\infty}^{\infty} P_\xi(\xi)\log(P_\xi(\xi))\,\mathrm{d}\xi\right]\mathrm{d}x \\ &= -\int_{-\infty}^{\infty} P_X(x)H(\xi)\,\mathrm{d}x = -H(\xi) \end{aligned}$$

and from part (a) we know that $H(\xi) = \frac{1}{2}\left[1 + \log\left(2\pi\sigma_\xi^2\right)\right]$, so

$$\begin{aligned} I(X, Y) &= H(Y) - H(Y|X) \\ &= \frac{1}{2}\log\left(\frac{2\pi(\sigma_X^2 + \sigma_\xi^2)}{2\pi\sigma_\xi^2}\right) \\ &= \frac{1}{2}\log\left(\frac{\sigma_X^2 + \sigma_\xi^2}{\sigma_\xi^2}\right) \\ &= \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_\xi^2}\right) \end{aligned}$$

- If σ_X is large and σ_ξ is small, then $I(X, Y)$ is large.
- If σ_X is small and σ_ξ is large, then $I(X, Y)$ is small or zero

4 (a) Given

$$m(n) = \frac{m(n+1)}{Nb_i} \implies m(n-1) = \frac{m(n)}{Nb_i}$$

the expected value of $x = \log(m(n))$ is

$$\mathrm{E}[x] = \sum_i x_i p_i$$

where the probability of horse i wins is p_i, so

$$\begin{aligned} \mathrm{E}[x] &= \sum_i (\log(m(n-1)) + \log(Nb_i))p_i \\ &= \sum_i \log(m(n-1))p_i + \sum_i \log N p_i + \sum_i \log(b_i)p_i \\ &= \sum_i \log(m(n-2)Nb_i)p_i + \sum_i \log N p_i + \sum_i \log(b_i)p_i \end{aligned}$$

which is a recursive structure so we get

$$\begin{aligned} &= \log(m(0)) + n(\mathrm{E}[\log N] + \mathrm{E}[\log b_i]) \\ &= \log(m(0)) + n\log N + n\,\mathrm{E}[\log b_i] \end{aligned}$$

(b) Finding where the derivative is zero and since we only care about the case of maximizing

$$\mathrm{E}[\log b_i] = \sum_i p_i \log(b_i)$$

and since b_i is a normalized vector:

$$\sum_i b_i = 1$$

we can use the method of Lagrange multipliers to find the maximum of $\mathrm{E}[\log b_i]$ (from Cover & Thomas Chapter 6):

$$\mathcal{L} = \sum_i p_i \log(b_i) + \lambda \sum_i b_i$$

and differentiating with respect to b_i:

$$\frac{\partial \mathcal{L}}{\partial b_i} = \frac{p_i}{b_i} + \lambda = 0 \implies b_i = -\frac{p_i}{\lambda}$$

and from the constraint:

$$\sum_i b_i = -\frac{\sum_i p_i}{\lambda} \implies \lambda = -1 \implies b_i = p_i$$

and to find the growth rate we can use

$$\mathrm{E}[x] = \log(m(0)) + \lambda_{\max}$$

where $n = 1$ is the expected growth rate after one race:

$$\begin{aligned}
\lambda_{\max} &= \log N + \sum_i p_i \log(p_i) \\
&= \log N - \sum_i p_i \log\left(\frac{1}{p_i}\right) \\
&= \log N - H(p)
\end{aligned}$$

(c) The max capital minus our capital based on our bet $b_i = q_i$ is

$$\begin{aligned}
\log(m(n)) &= \log(m(0)) + n \log N + n \sum_i p_i \log\left(p_i \frac{q_i}{p_i}\right) \\
&= \log(m(0)) + n \log N + n \sum_i p_i \log(p_i) + n \sum_i p_i \log\left(\frac{q_i}{p_i}\right) \\
&= \log(m(0)) + n \log N - nH(p) - nD_{KL}(p||q)
\end{aligned}$$

and since the maximimum capital is

$$m_{\max}(n) = m(0) e^{n \log N - nH(p)}$$

we can see the capital falls off exponentially

$$\begin{aligned}
m(n) &= m(0) e^{n \log N - nH(p) - nD_{KL}(p||q)} \\
&= m_{\max}(n) e^{-nD_{KL}(p||q)}
\end{aligned}$$

(d) Given

$$m(n+1) = \frac{1}{p_i} b_i m(n) \implies m(n) = \frac{1}{p_i} b_i m(n-1)$$

so the expected value of $\log(m(n))$:

$$\mathrm{E}[\log(m(n))] = \sum_i p_i \left(\log(m(n-1)) + \log\left(\frac{b_i}{p_i}\right)\right)$$

and since $b_i = p_i$ for the optimal strategy:

$$\begin{aligned}
\mathrm{E}[\log(m(n))] &= \mathrm{E}[\log(m(n-1))] + \sum_i p_i \log(1) \\
&= \mathrm{E}[\log(m(n-2))] + \sum_i p_i \log(1) + \sum_i p_i \log(1) \\
\dots \\
&= \log(m(0)) + n \log(1) \\
&= \log(m(0))
\end{aligned}$$

so we end up with the same money we started with! Therefore the long term capital growth is 0.
(e) (f) :(