(1) The joint probability distribution table:

(a)

$P(V,T)$	T					$P(V)$
V ☼	1/16	1/16	1/16	1/16	sum rows →	1/4
cloudy/dry	1/16	1/8	1/32	1/32		1/4
cloudy/rain	1/8	1/16	1/32	1/32		1/4
cloudy/snow	1/4	0	0	0		1/4
↓ sum columns						
$P(T)$	1/2	1/4	1/8	1/8		

$P(V)$: Entropy $H(V) = 2$ bits

$P(T)$: Entropy $= 7/4$ bits

$$\text{Joint entropy} = -\sum_{v,t} P(V=v, T=t) \log_2 P(V=v, T=t) = 27/8 \text{ bits.}$$

(b) The conditional distributions: (= rows, renormalized to sum to 1)

		Entropy
For V= "sunny"	$P(T \mid V=\text{sunny}) = \{1/4, 1/4, 1/4, 1/4\} \Rightarrow$	2 bits
	$P(T \mid V=\text{cloudy \& dry}) = \{1/4, 1/2, 1/8, 1/8\} \Rightarrow$	7/4 bits
	$P(T \mid V=\text{cloudy \& rain}) = \{1/2, 1/4, 1/8, 1/8\} \Rightarrow$	7/4 bits
	$P(T \mid V=\text{cloudy \& snow}) = \{1, 0, 0, 0\} \Rightarrow$	0 bits

So learning that V=sunny actually <u>increases</u> our uncertainty about T.

(c) But <u>on average</u>, $H(T|V) = \dfrac{2 + 7/4 + 7/4 + 0}{4} = \dfrac{11}{8}$ bits $< H(T) = \dfrac{7}{4}$

$P(V=v) = 1/4$ for all v

(d) Similarly: $P(V \mid T = \text{Miserably cold}) = \begin{pmatrix} 1/8 \\ 1/8 \\ 1/4 \\ 1/2 \end{pmatrix}$ $H = 7/4$

(renormalized colums)

$P(V \mid T = \text{Very cold}) = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \\ 0 \end{pmatrix}$ $H = 3/2$

$P(V \mid T = \text{Cold}) = \begin{pmatrix} 1/2 \\ 1/4 \\ 1/4 \\ 0 \end{pmatrix}$ $H = 3/2$

$P(V \mid T = \text{Chilly}) = \begin{pmatrix} 1/2 \\ 1/4 \\ 1/4 \\ 0 \end{pmatrix}$ $H = 3/2$

$\Rightarrow H(V|T) =$
$1/2 \cdot 7/4 + 1/4 \cdot 3/2 + 2 \cdot 1/8 \cdot 3/2$
$= 13/8$ bits

(e) $I(V,T) = H(V) - H(V|T) = 2 - 13/8 = 3/8$ bits.

Reassuringly: $H(T) - H(T|V) = 7/4 - 11/8 =$ also $3/8$ bits.

② (a) Applying the definition: $\leftarrow \int_{\mathbb{R}} \equiv \int_{-\infty}^{+\infty}$

$$H\left(P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\right) \equiv -\int_{\mathbb{R}} P(x) \ln P(x)\, dx$$

$$= -\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \cdot \left\{ -\frac{x^2}{2\sigma^2} + \ln \frac{1}{\sqrt{2\pi\sigma^2}} \right\} dx$$

$$= \ln\sqrt{2\pi\sigma^2} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx}_{1} + \frac{1}{2\sigma^2} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} x^2\, dx}_{\sigma^2}$$

$$= \frac{1}{2} + \ln\sqrt{2\pi\sigma^2} = \ln\sqrt{2\pi e\sigma^2}$$

(which can easily be negative if σ is small enough).

(b) For independent random variables, variances add.

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

(c) Consider a discrete version of this problem.

Let X take values $\{1,2,3\}$ and ξ take values $\{5,6,7\}$

What is $P(Y \equiv X + \xi$ takes value $8) = ?$

$$P(Y=8) = P(X=1, \xi=7) + P(X=2, \xi=6) + P(X=3, \xi=5)$$

$$= P_X(1)\cdot P_\xi(7) + P_X(2)\cdot P_\xi(6) + P_X(3)\cdot P_\xi(5)$$

$$= \sum_z P_\xi(z)\cdot P_X(8-z)$$

This problem is the continuous analog. $P_Y(y) = \int dz\, P_\xi(z)\, P_X(y-z)$.

(d) Specifically for Gaussians: let's take the integral in (c).

Gaussian integrals are a useful skill. Here is how it works.

$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$ is a normalized probability distribution

$$\Rightarrow \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$$

Let's write this as: $\int_{-\infty}^{+\infty} e^{-\frac{ax^2}{2}} dx = \sqrt{\frac{2\pi}{a}}$. (for any $a>0$).

Now let's compute $\int_{-\infty}^{+\infty} e^{-\frac{ax^2}{2}+bx+c} dx$

$$(\text{"complete the square"}) = \int_{-\infty}^{+\infty} e^{-\frac{a}{2}\left(x^2-\frac{2b}{a}x+\frac{b^2}{a^2}\right)+\frac{b^2}{2a}+c} dx$$

$$= e^{b^2/2a+c}\int_{-\infty}^{+\infty} e^{-\frac{a}{2}\left(x-\frac{b}{a}\right)^2} dx$$

$$(\tilde{x} \equiv x - b/a) \qquad = e^{b^2/2a+c}\int_{-\infty}^{+\infty} e^{-\frac{a}{2}\tilde{x}^2} d\tilde{x} = \sqrt{\frac{2\pi}{a}}\, e^{\frac{b^2}{2a}+c}.$$

In our case:

$$P_Y(y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(y-z)^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_\xi^2}} e^{-\frac{z^2}{2\sigma_\xi^2}} dz$$

$$= \frac{1}{2\pi\sigma_x\sigma_\xi}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{y^2-2yz+z^2}{\sigma_x^2}+\frac{z^2}{\sigma_\xi^2}\right)} dz$$

The exponent: $-\frac{1}{2}\left(\frac{y^2-2yz+z^2}{\sigma_x^2}+\frac{z^2}{\sigma_\xi^2}\right) = -\frac{az^2}{2}+bz+c$ with:

$$a = \frac{1}{\sigma_x^2}+\frac{1}{\sigma_\xi^2} = \frac{\sigma_\xi^2+\sigma_x^2}{\sigma_x^2\sigma_\xi^2} \qquad b = \frac{y}{\sigma_x^2} \qquad c = -\frac{y^2}{2\sigma_x^2}$$

Plugging this into the formula derived above:

$$P_Y(y) = \frac{1}{2\pi\sigma_x\sigma_\xi} \cdot \sqrt{\frac{2\pi}{a}}\, e^{b^2/2a + C}$$

$$= \frac{1}{2\pi\sigma_x\sigma_\xi}\sqrt{\frac{2\pi\sigma_x^2\sigma_\xi^2}{\sigma_\xi^2+\sigma_x^2}}\; e^{\frac{1}{2}\frac{y^2}{\sigma_x^4}\frac{\sigma_x^2\sigma_\xi^2}{\sigma_x^2+\sigma_\xi^2} - \frac{y^2}{2\sigma_x^2}}$$

$$= \frac{1}{\sqrt{2\pi(\sigma_\xi^2+\sigma_x^2)}}\; e^{-\frac{y^2}{2\sigma_x^2}\left(1-\frac{\sigma_\xi^2}{\sigma_x^2+\sigma_\xi^2}\right)} = \frac{1}{\sqrt{2\pi\sigma_Y^2}}\, e^{-\frac{y^2}{2\sigma_Y^2}}$$

with $\sigma_Y = \sqrt{\sigma_x^2+\sigma_\xi^2}$, exactly as expected. //

(e) $I(X,Y) = H(Y) - H(Y|X) \equiv H(Y) - \langle H(Y|X=x_0)\rangle_{x_0}$

Gaussian of width σ_Y

For any x_0, this distribution is a Gaussian of width σ_ξ

$$= \ln\sqrt{2\pi e\,\sigma_Y^2} - \ln\sqrt{2\pi e\,\sigma_\xi^2}$$

$$= \ln\frac{\sigma_Y}{\sigma_\xi} = \ln\frac{\sqrt{\sigma_x^2+\sigma_\xi^2}}{\sigma_\xi} = \ln\sqrt{1+\left(\frac{\sigma_x}{\sigma_\xi}\right)^2}$$

As $\sigma_\xi \to \infty$, $I(X,Y) \to 0$ (an infinitely noisy measurement $\Rightarrow$ no information).

If $\sigma_\xi \ll \sigma_x$: $I(X,Y) \sim \ln\frac{\sigma_x}{\sigma_\xi}$

3 (a) $\log m(n) = \log\left(N b_{i_n} m(n-1)\right) = \log m(n-1) + \log(N b_{i_n})$

$$= \log m(n-2) + \log(N b_{i_{n-1}}) + \log(N b_{i_n}) = \dots$$

$$= \log m_0 + \sum_{k \in \text{races}} \log(N \cdot b_{i_k})$$ where i_k = index of the horse that won the race #k.

initial capital

So over many races n, horse 1 wins $p_1 \cdot n$ times, etc., and we find $\log m(n) = \log m_0 + n \sum_i p_i \log(Nb_i)$

(b) We seek to maximize $\sum_i p_i \log(Nb_i)$ over all possible choices of b_i subject to condition $\sum_i b_i = 1$.

Let's change our bets by small amounts: $b_i \to b_i + \delta b_i$

The extremum condition requires that $\sum_i p_i \log(Nb_i)$ does not change. (To maximize $f(x)$ we require that $\left.\frac{d}{dx}\right|_{x=\text{optimum}} f(x) = 0$)

$$\sum_i p_i \log(N(b_i + \delta b_i)) - \sum_i p_i \log(Nb_i) = \sum_i p_i \log\left(1 + \frac{\delta b_i}{b_i}\right) \cong \sum_i \frac{p_i\, \delta b_i}{b_i}$$

This must be zero for any δb_i such that $\sum_i \delta b_i = 0$.

Take $\delta b_i = \{\epsilon, -\epsilon, 0, 0 \ldots 0\} \Rightarrow \frac{p_1}{b_1} = \frac{p_2}{b_2}$

$\Rightarrow$ The optimal b_i must satisfy $\frac{p_i}{b_i} = \underbrace{\text{const.}}_{\text{must be 1 to satisfy } \sum_i b_i = 1} \Rightarrow \underline{\underline{b_i = p_i}}$

For this optimal strategy, the growth rate of your capital is

$$\sum_i p_i \log Nb_i = \log N + \sum_i p_i \log p_i = \log N - H(p).$$

(c) If $b_i = q_i$ (the would-be optimum if probabilities were q_i) our capital grows as $\sum_i p_i \log(Nq_i)$ instead of $\sum_i p_i \log(Np_i)$ (the optimum). $\Rightarrow$ ~~We are losing out by: $\sum p_i \log$~~

$$m(n) \cong m_0\, e^{n\sum_i p_i \log(Nq_i)} = \underline{m_0\, e}^{\,n\sum_i p_i \log(Np_i) + n\{\sum_i p_i \log q_i/p_i\}} =$$

$$= m_{max}(n)\, e^{-n\sum_i\left(-p_i \log \frac{q_i}{p_i}\right)} \equiv m_{max}(n)\, e^{-n D_{KL}(p\|q)}.$$

(d) With the adjusted odds, long-term capital growth is

$$\sum_i p_i \log\Big(\underbrace{\tfrac{1}{p_i}}_{\text{previously } N} \cdot b_i\Big) \text{ maximized at } b_i = p_i \Rightarrow 0.$$

(e) Just as in (a), if race k is won by horse i_k:

$$\log m(n) = \log m(n-1) + \log\left(\frac{1}{p_i} \cdot b_{i_n}\right)$$

multiplication coefficient that comes from the bookmakers' payout upon our win. ~~we does not~~ The payout structure does not change from race to race.

But our bets now change from race to race: adjusted each time according to the information H_n we received. So what is our expected win? For a particular race #k, given the information we received, horse i wins with probability $p(i \text{ wins} | H_k)$

$$\Rightarrow \langle \log m(n)\rangle = \langle \log m(n-1)\rangle + \sum_i p(i \text{ wins}|H_n) \cdot \log\left(\frac{b_i}{p_i}\right)$$

And we are using $b_i = p(i \text{ wins} | H_n)$ ← the optimal strategy, given the information we have.

$$\Rightarrow \langle \log m(n)\rangle = \langle \log m(n-1)\rangle + \sum_i p(i \text{ wins}|H_n) \log \frac{p(i \text{ wins}|H_n)}{p_i}$$

$$\Rightarrow \langle \log m(n)\rangle = \log m_0 + \left\langle \sum_i -p(i \text{ wins}|H_n) \log \frac{p_i}{p(i \text{ wins}|H_n)} \right\rangle_{\text{races}}$$

$$= \log m_0 + \left\langle D_{KL}\big(P(i \text{ wins}|H_n) \,\|\, P(i \text{ wins})\big)\right\rangle_{n}$$

HW7 - solutions

$$\left\langle D_{KL}\left(P(X|Y) \,\|\, P(X)\right)\right\rangle_Y$$

$$= \left\langle -\sum_x P(X=x\,|\,Y=y) \log \frac{P(X=x)}{P(X=x\,|\,Y=y)} \right\rangle_y$$

$$= \sum_y P(Y=y) \cdot \left\{ -\sum_x P(X=x\,|\,Y=y) \log \frac{P(X=x)}{P(X=x\,|\,Y=y)} \right\}$$

$$= -\sum_{x,y} P(X=x\,|\,Y=y)\,P(Y=y) \log \frac{P(X=x)\cdot P(Y=y)}{P(X=x\,|\,Y=y)\,P(Y=y)}$$

conditional

joint

$$= -\sum_{x,y} P(X=x, Y=y) \log \frac{P(X=x)\cdot P(Y=y)}{P(X=x, Y=y)}$$

$$\equiv I(X,Y)$$