[redacted]

Problem 1 Let $X \sim Geo(p)$.

Then X has PMF $p(x) = \begin{cases} (1-p)^{x-1} p, & x \geq 1. \\ 0, & \text{otherwise.} \end{cases}$

We first consider $p \in (0,1)$.

The entropy of X is then given by

$$H(X) = \sum_{x=1}^{\infty} p(x) \log \frac{1}{p(x)}$$

$$= \sum_{x=1}^{\infty} (1-p)^{x-1} p \log \frac{1}{(1-p)^{x-1} p}$$

$$= \sum_{x=1}^{\infty} (1-p)^{x-1} p \left((x-1) \log \frac{1}{1-p} + \log \frac{1}{p} \right)$$

$$= p \log \frac{1}{1-p} \sum_{x=1}^{\infty} (1-p)^{x-1}(x-1) + p \log \frac{1}{p} \sum_{x=1}^{\infty} (1-p)^{x-1} \quad \cdots (1).$$

We separately calculate the two summations in (1):

① $\sum_{x=1}^{\infty} (1-p)^{x-1}(x-1) = \sum_{x=2}^{\infty} (1-p)^{x-1}(x-1)$

(1 Cont'd)

$$\overset{(u=x-1)}{=} \sum_{u=1}^{\infty} (1-p)^{u} u$$

$$= (1-p) \sum_{u=1}^{\infty} (1-p)^{u-1} u. \quad \cdots (2)$$

Recall that $\sum_{k=0}^{\infty} t^k = \frac{1}{1-t}$ for $t \in (-1,1) \quad \cdots (3)$

Differentiating on both sides of (3) yields

$$\sum_{k=1}^{\infty} k t^{k-1} = \frac{1}{(1-t)^2} \text{ for } t \in (-1,1) \quad \cdots (4).$$

Since $p \in (0,1)$ by assumption, we have $(1-p) \in (0,1) \subset (-1,1)$ as well. Then we can put $t = 1-p$ in (4), which together with (2) gives

$$\sum_{x=1}^{\infty} (1-p)^{x-1}(x-1) = (1-p) \frac{1}{(1-(1-p))^2}$$

$$= \frac{1-p}{p^2}. \quad \cdots (5)$$

(1 cont'd) ② $\sum_{x=1}^{\infty} (1-p)^{x-1} \overset{u=x-1}{=\!=} \sum_{u=0}^{\infty} (1-p)^u$

$$\overset{\text{by (3)}}{=} \frac{1}{1-(1-p)} = \frac{1}{p} \quad \cdots (6)$$

Putting (5) & (6) into (1) yields

$$H(X) = p \log \frac{1}{1-p} \frac{1-p}{p^2} + p \log \frac{1}{p} \frac{1}{p}$$

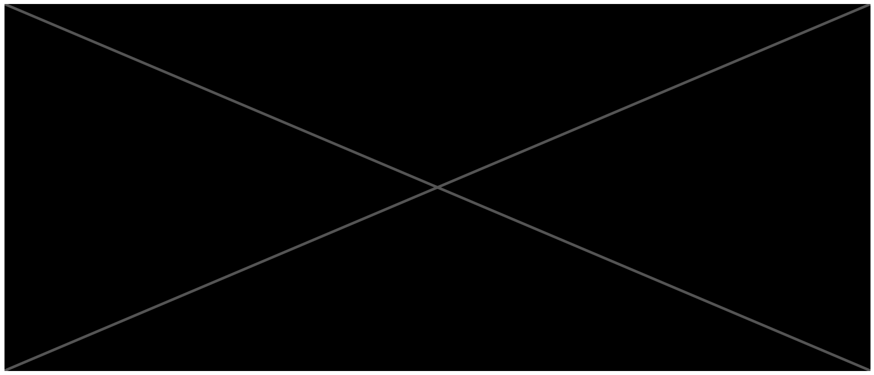
$$= \frac{1}{p} \left[(1-p) \log \frac{1}{1-p} + p \log \frac{1}{p} \right]$$

$$= \frac{1}{p} H(p) \quad \text{for } 0 < p < 1 \quad \cdots (7)$$

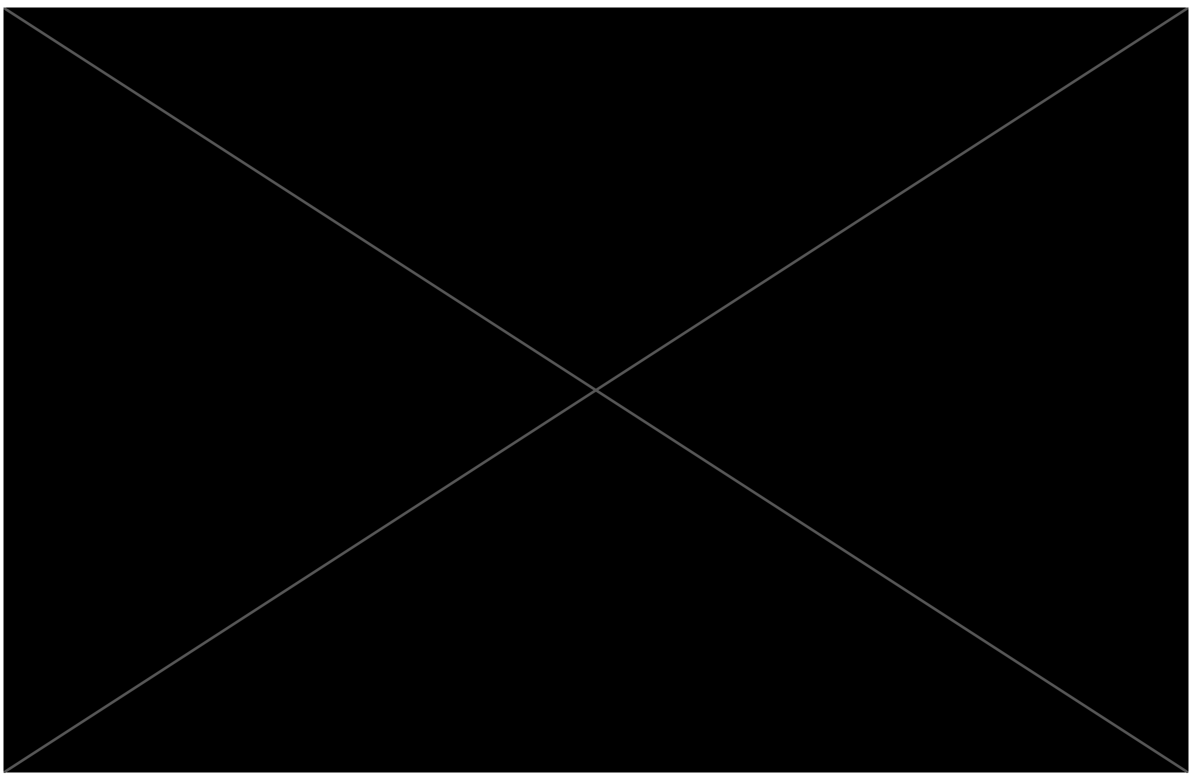
Now consider $p=1$. Then $X=1$ with probability 1. That is, X is deterministic, which has entropy $0 = \frac{1}{1} H(1)$.

Combining with (7), we have

$$\underline{H(X) = \frac{1}{p} H(p) \quad \text{for } 0 < p \leq 1} \quad \#$$



- **Problem 1.** Find the entropy of a geometric random variable X with parameter p. Next, assume that Y is the number of coin flips until the second head appears. Prove that $H(Y) \leqslant 2H(X)$ and explain intuitively why this is the case.



2. Because the geometric distribution is memoryless, we can say that $Y = X_1 + X_2$, as the first realization has no bearing on the distribution of the second realization, which is identical to the first distribution. As a result,

$$H(Y) = H(X_1 + X_2) \leq H(X_1, X_2) = H(X_1) + H(X_2) = 2H(X) \quad (10)$$

- **Problem 2.** An urn contains $\mathtt{b}$ blue, $\mathtt{g}$ green and $\mathtt{r}$ red balls. Does drawing two balls from the urn with replacement have higher entropy than drawing two balls without replacement? Justify your answer.

Let X_1 be the result of the first draw, and let X_2 be the result of the second draw. We know that

$$H(X_2, X_1) = H(X_1) + H(X_2|X_1)$$

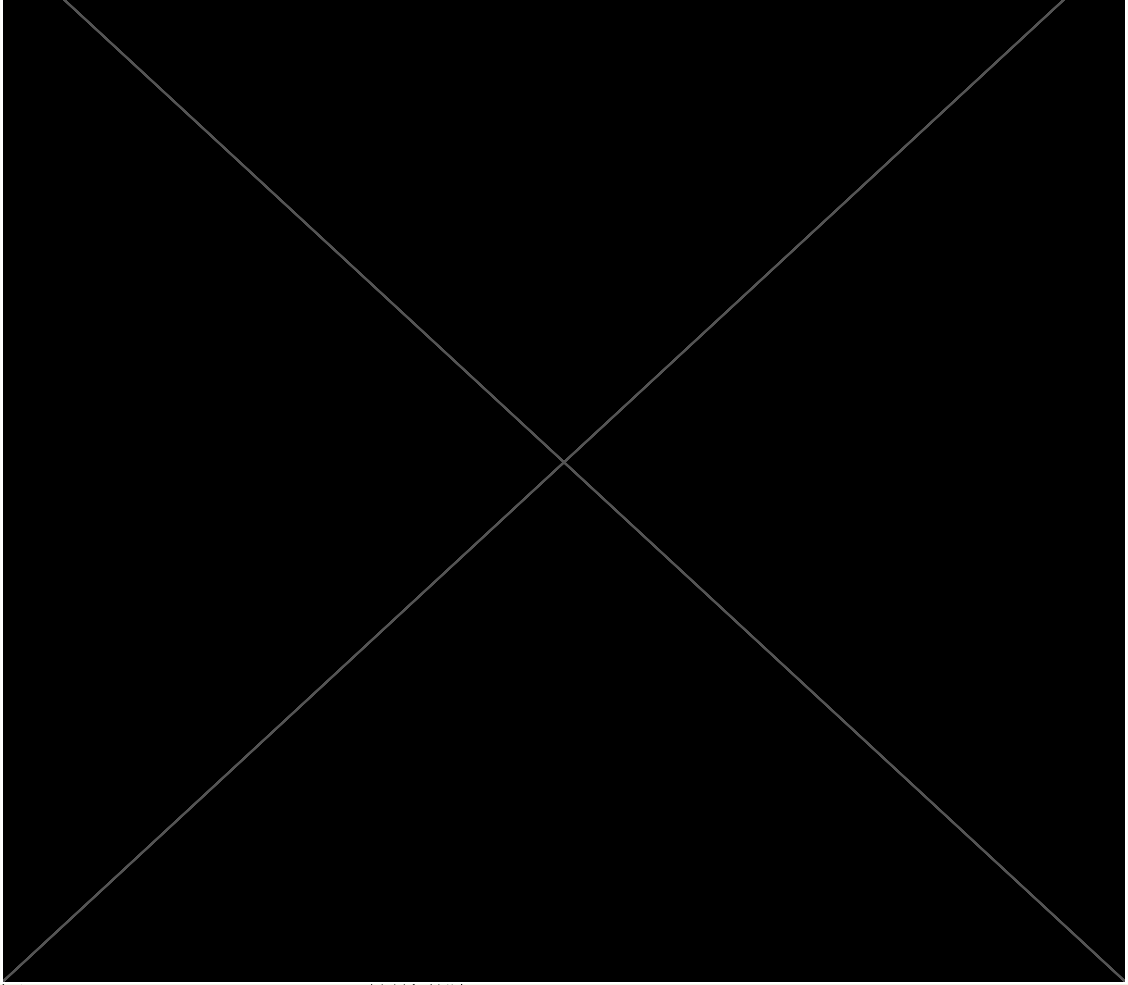
If the balls are being replaced, then the first draw has no bearing on the second draw. Therefore

$$H_{\text{replacement}}(X_2, X_1) = H(X_1) + H(X_2)$$

. If the balls are not being replaced, the draws are clearly not independent. Additionally, conditioning can only reduce entropy, in this case resulting in a strict inequality. So

$$H_{\text{no replacement}}(X_2|X_1) < H_{\text{replacement}}(X_2|X_1) = H(X_2)$$

.



Problem 3 (a) $\rho = 1 - \dfrac{H(X_2|X_1)}{H(X_1)}$

$$= \frac{H(X_1) - H(X_2|X_1)}{H(X_1)}$$

$$= \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} \quad \cdots (2.3)$$

$$= \frac{I(X_1; X_2)}{H(X_1)} \quad ,$$

where (23) holds since X_1 has the same distribution as X_2. QED.

(b) From (a) and the fact that $I(X_1;X_2) \geq 0$ we have

$$\rho = \frac{I(X_1;X_2)}{H(X_1)} \geq 0. \quad \cdots (24)$$

On the other hand, since $H(X_2|X_1) \geq 0$ we have

$$\rho \leq 1 - \frac{0}{H(X_1)} = 1. \quad \cdots (25)$$

Combining (24) & (25) gives $0 \leq \rho \leq 1$. QED.

(c) From (a) we have $\rho = \dfrac{I(X_1;X_2)}{H(X_1)} = 0$.

Assume $H(X_1) < \infty$.

$\Rightarrow I(X_1;X_2) = 0$

$\Rightarrow D(P_{X_1,X_2}(x_1,x_2) \,\|\, P_{X_1}(x_1) P_{X_2}(x_2)) = 0$

$\Rightarrow P_{X_1,X_2}(x_1,x_2) = P_{X_1}(x_1) P_{X_2}(x_2)$

$\Rightarrow X_1$ and X_2 are independent and thus i.i.d. #

(d) By definition, if

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} = 1,$$ then

$$\frac{H(X_2|X_1)}{H(X_1)} = 0$$

Assume $H(X_1) < \infty$ again.

$\Rightarrow H(X_2|X_1) = 0$

$\Rightarrow X_2$ is a deterministic function of X_1

Rmk If we take $H(X_1) = H(X_2) = \infty$ into consideration, then the answer to (c) should also include the case that $I(X_1; X_2) < \infty$ but $H(X_1) = \infty$. Similarly, the answer to (d) includes $H(X_2|X_1) < \infty$ but $H(X_1) = \infty$.

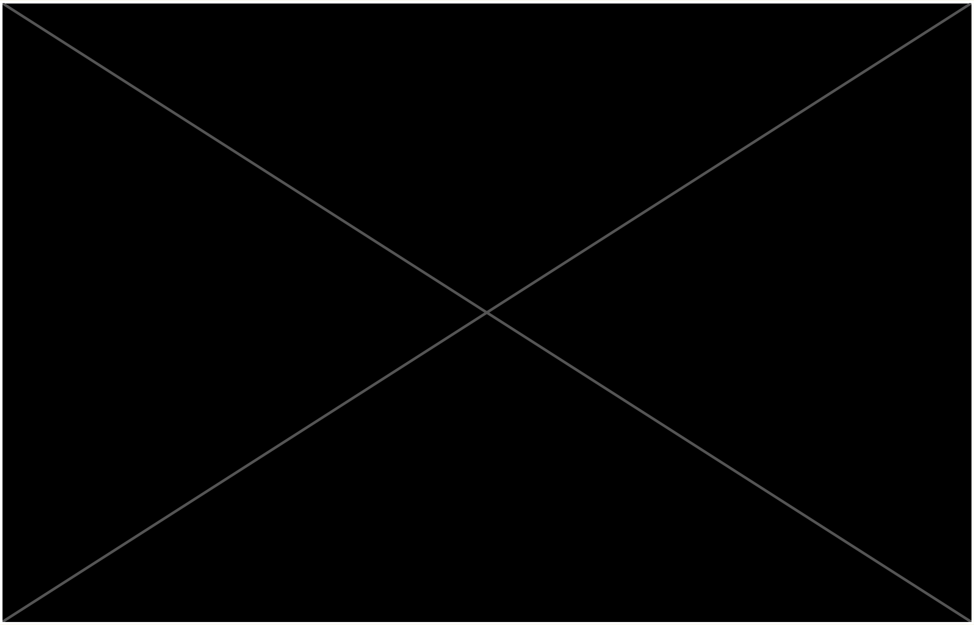
- **Problem 4.** Solve the following two problems:
  - Find the mutual information between the top and bottom side of a flipped fair coin.
  - Find the mutual information between the top and bottom side of a rolled fair die which has exactly five sides.

1. Let X be a bernoulli random variable with parameter 0.5 (indicator for heads). Similarly, let $Y = 1 - X$ (indicator for tails).

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) && (19)\\
&= H(X) && (20)\\
&= 1 && (21)
\end{aligned}
$$

2. I was not able to determine the "top" and "bottom" of a 5-sided die, since if one face is resting on a table, the opposite "face" is actually a corner. Instead, I solved the problem with a 6-sided die. A typical six-sided die has pairs 1-6, 3-4, 2-5. Similarly to the first part of this question, knowing the top side deterministically resolves the bottom side. Let X be the random variable that denotes the result of the top side, and let Y denote the bottom side. Then

$$I(X;Y) = H(X) = \log(6)$$



Problem 5 Consider a random variable X over the space

$$\mathcal{F} = \{ x_i : 1 \le i \le M \} \text{ such that}$$

$$P (X = x_i) = w_i \text{ for each } 1 \le i \le M.$$

Define $f(x) = \frac{1}{x}$ on $x \in (0, \infty)$.

Note that $f''(x) = \frac{2}{x^3} > 0$ on $(0, \infty)$, and thus

f is convex on $(0, \infty)$.

Since $x_i > 0$ for each $1 \le i \le M$, we can apply Jensen's inequality to have

$$\mathbb{E}[f(X)] \ge f(\mathbb{E}[X]). \quad \cdots (26)$$

Note that $\mathbb{E}[f(X)] = \mathbb{E}\left[\frac{1}{X}\right] = \sum_{i=1}^{M} \frac{w_i}{x_i}$, and that

$$f(\mathbb{E}[X]) = f\left(\sum_{i=1}^{M} w_i x_i\right) = \frac{1}{\sum_{i=1}^{M} w_i x_i}.$$

Therefore, (26) becomes

$$\sum_{i=1}^{M} \frac{w_i}{x_i} \ge \frac{1}{\sum_{i=1}^{M} w_i x_i}.$$

Q.E.D.

Problem 7 Since

1. We shall first find these values:

$$H_0 = \frac{1}{1-0} \log \sum_{i=1}^{M} P_i^0 = \underline{\log M} \cdots (28) \quad \#$$

$$H_1 = \lim_{\alpha \to 1} \frac{1}{1-\alpha} \log \sum_{i=1}^{M} P_i^\alpha$$

$$\overset{\text{L'Hopital}}{=} \lim_{\alpha \to 1} \frac{\frac{1}{\ln 2} \frac{1}{\sum_{i=1}^{M} P_i^\alpha} \sum_{i=1}^{M} P_i^\alpha \ln P_i}{-1}$$

$$= \frac{-1}{\sum_{i=1}^{M} P_i} \sum_{i=1}^{M} P_i \log P_i$$

$$= \underline{\sum_{i=1}^{M} P_i \log \frac{1}{P_i} = H} \cdots (29) \quad \#$$

(Shannon's entropy).

$$H_2 = \frac{1}{1-2} \log \sum_{i=1}^{n} p_i^2$$

$$= -\log \sum_{i=1}^{n} p_i^2 \cdots (30)$$

$\#$

$$H_\infty = \lim_{\alpha \to \infty} \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha$$

L'Hopital

$$= \lim_{\alpha \to \infty} \frac{\frac{1}{\ln 2} \frac{1}{\sum_{i=1}^{n} p_i^\alpha} \sum_{i=1}^{n} p_i^\alpha \ln p_i}{-1}$$

$$= -\lim_{\alpha \to \infty} \frac{1}{\sum_{i=1}^{n} p_i^\alpha} \sum_{i=1}^{n} p_i^\alpha \log p_i \cdots (31)$$

Define $p_{max} = \max_{1 \le i \le n} p_i$, $T = \{ i \in \{1, \cdots, n\} \mid p_i = p_{max} \}$.

Then (31) becomes

$$H_\infty = -\lim_{\alpha\to\infty} \frac{P_{max} \sum_{i=1}^{m} \left(\frac{P_i}{P_{max}}\right)^\alpha \log P_i}{P_{max} \sum_{i=1}^{m} \left(\frac{P_i}{P_{max}}\right)^\alpha}$$

$$= -\lim_{\alpha\to\infty} \frac{\sum_{i=1}^{m} \left(\frac{P_i}{P_{max}}\right)^\alpha \log P_i}{\sum_{i=1}^{m} \left(\frac{P_i}{P_{max}}\right)^\alpha}$$

$$= -\frac{\sum_{i\in T} 1 \log P_i}{\sum_{i\in T} 1}$$

$$= \frac{-T \log P_{max}}{T} = \underline{-\log P_{max}} \ \#$$

① $H_0 \geq H_1$:

$H_1(X) = H(X) \leq \log(|\mathrm{supp}(X)|) = \log M = H_0(X)$,

where the inequality has been proven by class via the fact that $D(X \,\|\, \mathrm{Unif}\{1, \dots, M\}) \geq 0$. QED.

② $H_1 \geq H_2$:

Consider a discrete RV Y on space $\mathcal{F} = \{p_i \mid i = 1 \sim M\}$ with $P(Y = p_i) = p_i$ for $i = 1 \sim M$.

Since $t \mapsto -\log t$ is convex, we have from Jensen's inequality that

$$H_2(X) = -\log \sum_{i=1}^{M} p_i^2$$

$$= -\log \sum_{i=1}^{n} p_i P(Y = p_i)$$

$$= -\log \mathbb{E}[Y]$$

$$\leq \mathbb{E}[-\log Y]$$

$$= \sum_{i=1}^{M} P(Y = p_i)(-\log p_i)$$

$$= -\sum_{i=1}^{n} P_i \log P_i = H_1(X).$$

QED.

③ $H_2 \geq H_\infty$:

Note that $P_i \leq P_{max}$ for $i = 1 \sim M$ and that $\log$ is a monotone non-decreasing function.

Thus,

$$H_2 = -\log \sum P_i^2 \geq H_\infty = -\log P_{max}$$

iff

$$\sum P_i^2 \leq P_{max},$$

which is clearly true since

$$\sum P_i^2 \leq \sum P_i P_{max} = P_{max} \sum P_i = P_{max}.$$

(Recall that P_i is a distribution). QED.

2. ① $H_\alpha(X) \geq 0$:

Discuss on three cases:

$\boxed{1}$ If $0 \leq \alpha < 1$, then $P_i^\alpha \geq P_i$ for $i = 1 \sim M$.

In addition, $1 - \alpha > 0$. Thus,

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum p_i^\alpha$$

$$\geq \frac{1}{1-\alpha} \log \sum p_i = \log 1 = 0$$

[2] $\alpha = 1$: $H_1(X) = H(X) \geq 0$ by class.

[3] $\alpha > 1$: $p_i^\alpha \leq p_i$ for $i = 1 \sim M$. In addition,

$1 - \alpha < 0$. Thus

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum p_i^\alpha$$

$$\geq \log p_i = \log 1 = 0.$$

(2) H_α is cvx. for $\alpha \leq 1$:

The case $\alpha = 1$ has been proven in class.

Now consider X_1 and X_2 being two RVs w. pmf. p_i and q_i, respectively.

Note that since $\alpha < 1$, $t \mapsto t^\alpha$ is a concave function. Together with the fact that log is concave, we have for each $\beta \in (0,1)$ that

$$H_\alpha(\beta X_1 + (1-\beta) X_2)$$

$$= \frac{1}{1-\alpha} \log \left[\sum_{i=1}^{n} (\beta p_i + (1-\beta) \xi_i)^\alpha \right]$$

$$\geq \frac{1}{1-\alpha} \log \sum_{i=1}^{n} \left(\beta p_i^\alpha + (1-\beta) \xi_i^\alpha \right) \quad \cdots (32)$$

$$= \frac{1}{1-\alpha} \log \left(\beta \sum_{i=1}^{n} p_i^\alpha + (1-\beta) \sum_{i=1}^{n} \xi_i^\alpha \right)$$

$$\geq \frac{1}{1-\alpha} \left[\beta \log \sum_{i=1}^{n} p_i^\alpha + (1-\beta) \log \sum_{i=1}^{n} \xi_i^\alpha \right] \cdots (33)$$

$$= \beta H_\alpha(X_1) + (1-\beta) H_\alpha(X_2),$$

where (32) holds by Jensen's inequality and the monoticity of log, and (33) holds by Jensen's inequality again. QED.