Problem Set 1 — *Due Friday, October 12, before class starts*

For the Exercise Sessions on Sep 28 and Oct 5

Last name	First name	SCIPER Nr	Points

Problem 1: Divergence and L_1

Suppose p and q are two probability mass functions on a finite set $\mathcal{U}$. (I.e., for all $u \in \mathcal{U}$, $p(u) \geq 0$ and $\sum_{u\in\mathcal{U}} p(u) = 1$; similarly for q.)

(a) Show that the L_1 distance $\|p - q\|_1 := \sum_{u\in\mathcal{U}} |p(u) - q(u)|$ between p and q satisfies

$$\|p - q\|_1 = 2 \max_{\mathcal{S}:\mathcal{S}\subset\mathcal{U}} p(\mathcal{S}) - q(\mathcal{S})$$

with $p(\mathcal{S}) = \sum_{u\in\mathcal{S}} p(u)$ (and similarly for q), and the maximum is taken over all subsets $\mathcal{S}$ of $\mathcal{U}$.

For α and β in $[0,1]$, define the function $d_2(\alpha\|\beta) := \alpha \log \frac{\alpha}{\beta} + (1-\alpha)\log\frac{1-\alpha}{1-\beta}$. Note that $d_2(\alpha\|\beta)$ is the divergence of the distribution $(\alpha, 1-\alpha)$ from the distribution $(\beta, 1-\beta)$.

(b) Show that the first and second derivatives of d_2 with respect to its first argument α satisfy $d_2'(\beta\|\beta) = 0$ and $d_2''(\alpha\|\beta) = \frac{\log e}{\alpha(1-\alpha)} \geq 4 \log e$.

(c) By Taylor's theorem conclude that

$$d_2(\alpha\|\beta) \geq 2(\log e)(\alpha - \beta)^2.$$

(d) Show that for any $\mathcal{S} \subset \mathcal{U}$

$$D(p\|q) \geq d_2(p(\mathcal{S})\|q(\mathcal{S}))$$

[Hint: use the data processing theorem for divergence.]

(e) Combine (a), (c) and (d) to conclude that

$$D(p\|q) \geq \tfrac{\log e}{2}\|p - q\|_1^2.$$

(f) Show, by example, that $D(p\|q)$ can be $+\infty$ even when $\|p - q\|_1$ is arbitrarily small. [Hint: considering $\mathcal{U} = \{0, 1\}$ is sufficient.] Consequently, there is no generally valid inequality that upper bounds $D(p\|q)$ in terms of $\|p - q\|_1$.

Solution

(a) For any set $\mathcal{S}$, we have

$$p(\mathcal{S}) - q(\mathcal{S}) = \sum_{u\in\mathcal{S}} p(u) - q(u) \le \sum_{u\in\mathcal{S}} |p(u) - q(u)|. \tag{1}$$

Similarly for the compliment set of $\mathcal{S}$, we also have

$$q(\mathcal{S}^c) - p(\mathcal{S}^c) = \sum_{u\in\mathcal{S}^c} q(u) - p(u) \le \sum_{u\in\mathcal{S}^c} |p(u) - q(u)|. \tag{2}$$

Note that $p(\mathcal{S}) + p(\mathcal{S}^c) = q(\mathcal{S}) + q(\mathcal{S}^c) = 1$. Thus $q(\mathcal{S}^c) - p(\mathcal{S}^c) = p(\mathcal{S}) - q(\mathcal{S})$. Therefore, we have

$$2(p(\mathcal{S}) - q(\mathcal{S})) \le \sum_{u\in\mathcal{S}} |p(u) - q(u)| + \sum_{u\in\mathcal{S}^c} |p(u) - q(u)| = \sum_{u\in\mathcal{U}} |p(u) - q(u)| = \|p - q\|_1 \tag{3}$$

For the choice $\mathcal{S} = \{u : p(u) > q(u)\}$, we have

$$p(\mathcal{S}) - q(\mathcal{S}) = \sum_{u\in\mathcal{S}} p(u) - q(u) = \sum_{u\in\mathcal{S}} |p(u) - q(u)| \tag{4}$$

$$q(\mathcal{S}^c) - p(\mathcal{S}^c) = \sum_{u\in\mathcal{S}^c} q(u) - p(u) = \sum_{u\in\mathcal{S}^c} |p(u) - q(u)| \tag{5}$$

So, for this $\mathcal{S}$, we have $2(p(\mathcal{S}) - q(\mathcal{S})) = \|p - q\|_1$.

(b): Since $d_2(\alpha||\beta) = \alpha \log \frac{\alpha}{\beta} + (1-\alpha)\log\frac{1-\alpha}{1-\beta}$,

$$d_2'(\alpha||\beta) = \frac{\partial d_2(\alpha||\beta)}{\partial \alpha} = \log\frac{\alpha}{\beta} + \log e - \log\frac{1-\alpha}{1-\beta} - \log e = \log\frac{\alpha(1-\beta)}{\beta(1-\alpha)} \tag{6}$$

Therefore, we have $d_2'(\beta||\beta) = 0$.

$$d_2''(\alpha||\beta) = \frac{\log e}{\alpha(1-\alpha)} \ge 4\log e \tag{7}$$

where equality achieves when $\alpha = 1/2$.

(c): Taylor theorem says that for any f for which f'' is continuous

$$f(\alpha) = f(\beta) + (\alpha - \beta)f'(\beta) + (1/2)(\alpha-\beta)^2 f''(x_i) \tag{8}$$

where x_i is a value between α and β. With $f(\alpha) = d_2(\alpha\|\beta)$, we thus have

$$d_2(\alpha\|\beta) = 0 + 0 + (1/2)(\alpha-\beta)^2 f''(x_i) \ge 2\log(e)(\alpha-\beta)^2 \tag{9}$$

(d) Consider a deterministic channel with binary output

$$V = \begin{cases} 1, & \text{if } V \in \mathcal{S} \\ 0, & \text{if } V \notin \mathcal{S} \end{cases} \tag{10}$$

Thus,

$$\begin{aligned} d_2(p(\mathcal{S})\|q(\mathcal{S})) &= p(\mathcal{S})\log\frac{p(\mathcal{S})}{q(\mathcal{S})} + (1-p(\mathcal{S}))\log\frac{1-p(\mathcal{S})}{1-q(\mathcal{S})} && (11)\\ &= p(V=1)\log\frac{p(V=1)}{q(V=1)} + p(V=0)\log\frac{p(V=0)}{q(V=0)} && (12)\\ &= D(p_V\|q_V) && (13)\end{aligned}$$

By data processing theorem for divergence, $D(p\|q) \geq D(p_V\|q_V)$

(e) Combine (a),(c) and (d) and choosing $\mathcal{S} = \{u : p(u) > q(u)\}$, we have $\forall \mathcal{S}$

$$D(p\|q) \geq d_2(p(\mathcal{S})\|q(\mathcal{S})) \geq 2(\log e)(p(\mathcal{S}) - q(\mathcal{S}))^2 = \frac{\log e}{2}\|p - q\|_1^2 \tag{14}$$

(f) Let p be Bernoulli distribution with probability ϵ to be 1 and q is 0 with probability 1. Then

$$D(p\|q) = p(1)\log\frac{p(1)}{q(1)} + p(0)\log\frac{p(0)}{q(0)} = +\infty \tag{15}$$

But $\|p - q\|_1 = 2\epsilon$.

Problem 2: Other Divergences

Suppose f is a convex function defined on $(0, \infty)$ with $f(1) = 0$. Define the f-divergence of a distribution p from a distribution q as

$$D_f(p\|q) := \sum_u q(u)f(p(u)/q(u)).$$

In the sum above we take $f(0) := \lim_{t\to 0} f(t)$, $0f(0/0) := 0$, and $0f(a/0) := \lim_{t\to 0} tf(a/t) = a\lim_{t\to 0} tf(1/t)$.

(a) Show that for any non-negative a_1, a_2, b_1, b_2 and with $A = a_1 + a_2$, $B = b_1 + b_2$,

$$b_1 f(a_1/b_1) + b_2 f(a_2/b_2) \geq Bf(A/B);$$

and that in general, for any non-negative $a_1, \ldots, a_k$, $b_1, \ldots, b_k$, and $A = \sum_i a_i$, $B = \sum_i b_i$, we have

$$\sum_i b_i f(a_i/b_i) \geq Bf(A/B).$$

[Hint: since f is convex, for any $\lambda \in [0,1]$ and any $x_1, x_2 > 0$ $\lambda f(x_1) + (1-\lambda)f(x_2) \geq f(\lambda x_1 + (1-\lambda)x_2)$; consider $\lambda = b_1/B$.]

(b) Show that $D_f(p\|q) \geq 0$.

(c) Show that D_f satisfies the data processing inequality: for any transition probability kernel $W(v|u)$ from $\mathcal{U}$ to $\mathcal{V}$, and any two distributions p and q on $\mathcal{U}$

$$D_f(p\|q) \geq D_f(\tilde{p}\|\tilde{q})$$

where $\tilde{p}$ and $\tilde{q}$ are probability distributions on $\mathcal{V}$ defined via $\tilde{p}(v) := \sum_u W(v|u)p(u)$, and $\tilde{q}(v) := \sum_u W(v|u)q(u)$,

(d) Show that each of the following are f-divergences.

i. $D(p\|q) := \sum_u p(u)\log(p(u)/q(u))$. [Warning: $\log$ is not the right choice for f.]

ii. $R(p\|q) := D(q\|p)$.

iii. $1 - \sum_u \sqrt{p(u)q(u)}$

iv. $\|p - q\|_1$.

v. $\sum_u (p(u) - q(u))^2/q(u)$

Solution

(a) Since f is convex, for any $\lambda \in [0,1]$ and any $x_1, x_2 >$ we have

$$\lambda f(x_1) + (1-\lambda) f(x_2) \geq f(\lambda x_1 + (1-\lambda) x_2) \tag{16}$$

By substitution $x_1 = a_1/b_1$, $x_2 = a_2/b_2$ and $\lambda = b_1/(b_1+b_2)$:

$$\frac{b_1}{b_1+b_2} f(\frac{a_1}{b_1}) + (1 - \frac{b_1}{b_1+b_2}) f(\frac{a_2}{b_2}) \geq f(\frac{b_1}{b_1+b_2}\frac{a_1}{b_1} + (1 - \frac{b_1}{b_1+b_2})\frac{a_2}{b_2}) \tag{17}$$

$$\Leftrightarrow b_1 f(\frac{a_1}{b_1}) + b_2 f(\frac{a_2}{b_2}) \geq B f(A/B) \tag{18}$$

Let $A_k = \sum_{i=1}^k a_i$, $B_k = \sum_{i=1}^k b_i$. As we have proved that the following inequality holds for $k = 1, 2$:

$$\sum_{i=1}^k b_i f(a_i/b_i) \geq B_k f(A_k/B_k). \tag{19}$$

We assume that it also holds for $k = n$. For $k = n+1$, we have

$$\begin{aligned}
\sum_{i=1}^{n+1} b_i f(a_i/b_i) &= \sum_{i=1}^{n} b_i f(a_i/b_i) + b_{n+1} f(a_{n+1}/b_{n+1}) && (20)\\
&\geq B_n f(A_n/B_n) + b_{n+1} f(a_{n+1}/b_{n+1}) && (21)\\
&\geq B_{n+1} f(A_{n+1}/B_{n+1}) && (22)
\end{aligned}$$

By induction, for all any non-negative k, we have

$$\sum_{i=1}^k b_i f(a_i/b_i) \geq B_k f(A_k/B_k). \tag{23}$$

(b) $D_f(p\|q) = \sum_u q(u) f(p(u)/q(u)) \geq (\sum_u q(u)) f(\frac{\sum_u p(u)}{\sum_u q(u)}) = 1 f(1) = 0$.

(c)

$$\begin{aligned}
D_f(p\|q) = \sum_u q(u) f(p(u)/q(u)) &= \sum_u \sum_v W(v|u) q(u) f(p(u)/q(u)) && (24)\\
&= \sum_u \sum_v W(v|u) q(u) f(W(v|u)p(u)/(W(v|u)q(u))) && (25)\\
&\geq \sum_v (\sum_u W(v|u) q(u)) f\left(\frac{\sum_u W(v|u)p(u)}{\sum_u W(v|u)q(u)}\right) && (26)\\
&= \sum_v \tilde{q}(v) f(\tilde{p}(v)/\tilde{q}(v)) && (27)\\
&= D_f(\tilde{p}\|\tilde{q}) && (28)
\end{aligned}$$

(d)

i. $D(p\|q) := \sum_u p(u) \log(p(u)/q(u)) = \sum_u q(u) \frac{p(u)}{q(u)} \log \frac{p(u)}{q(u)}$. So $f(t) = t \log t$.

ii. $R(p\|q) := D(q\|p) = \sum_u p(u) \log(p(u)/q(u)) = \sum_u p(u)(-\log(q(u)/p(u)))$. So $f(t) = -\log t$.

iii. $1 - \sum_u \sqrt{p(u)q(u)} = \sum_u q(u) \left(1 - \sqrt{p(u)/q(u)}\right)$. So $f(t) = 1 - \sqrt{t}$.

iv. $\|p - q\|_1 = \sum_u |p(u) - q(u)| = \sum_u q(u) |p(u)/q(u) - 1|$. So $f(t) = |t - 1|$.

v. $\sum_u (p(u) - q(u))^2 / q(u) = \sum_u q(u)(p(u)/q(u) - 1)^2$. So $f(t) = (t-1)^2$.

Problem 3: Entropy and pairwise independence

Suppose X, Y, Z are pairwise independent fair flips, i.e., $I(X;Y) = I(Y;Z) = I(Z;X) = 0$.

(a) What is $H(X,Y)$?

(b) Give a lower bound to the value of $H(X,Y,Z)$.

(c) Give an example that achieves this bound.

Solution

(a) Since X, Y, Z are pairwise independent fair flips, $H(X) = H(Y) = H(Z) = 1$. $H(X,Y) = H(X) + H(Y|X) = H(X) + H(Y) - I(X;Y) = 2$.

(b) $H(X,Y,Z) = H(X,Y) + H(Z|X,Y) \geq H(X,Y) = 2$

(c) Let $Z = X + Y \mod 2$, then $H(Z|X,Y) = 0$ and $H(X,Y,Z) = H(X,Y)$.

Problem 4: Generating fair coin flips from biased coins

Suppose $X_1, X_2, \ldots$ are the outcomes of independent flips of a biased coin. Let $\Pr(X_i = 1) = p$, $\Pr(X_i = 0) = 1 - p$, with p unknown. By processing this sequence we would like to obtain a sequence $Z_1, Z_2, \ldots$ of *fair* coin flips.

Consider the following method: We process the X sequence in sucssive pairs, (X_1X_2), (X_3X_4), (X_5X_6), mapping (01) to 0, (10) to 1, and the other outcomes (00) and (11) to the empty string. After processing X_1, X_2, we will obtain either nothing, or a bit Z_1.

(a) Show that, if a bit is obtained, it is fair, i.e., $\Pr(Z_1 = 0) = \Pr(Z_1 = 1) = 1/2$.

In general we can process the X sequence in successive n-tuples via a function $f : \{0,1\}^n \to \{0,1\}^*$ where $\{0,1\}^*$ denote the set of all finite length binary sequences (including the empty string λ). [The case in (a) is the function $f(00) = f(11) = \lambda$, $f(01) = 0$, $f(10) = 1$. The function f is chosen such that $(Z_1, \ldots, Z_K) = f(X_1, \ldots, X_n)$ are i.i.d., and fair (here K may depend on $(X_1, \ldots, X_K)$.

(b) With $h_2(p) = -p \log p - (1-p) \log(1-p)$, prove the following chain of (in)equalities.

$$\begin{aligned} nh_2(p) &= H(X_1, \ldots, X_n) \\ &\geq H(Z_1, \ldots, Z_K, K) \\ &= H(K) + H(Z_1 \ldots, Z_K|K) \\ &= H(K) + E[K] \\ &\geq E[K]. \end{aligned}$$

Consequently, on the average no more than $nh_2(p)$ fair bits can be obtained from $(X_1, \ldots, X_n)$.

(c) Find a good f for $n = 4$.

Solution

(a) Since $\Pr(X_1 = 0, X_2 = 1) = \Pr(X_1 = 0)\Pr(X_2 = 1) = p(1-p)$ and $\Pr(X_1 = 1, X_2 = 0) = \Pr(X_1 = 1)\Pr(X_2 = 0) = p(1-p)$, the probability of $\Pr(Z_1 = 0) = \Pr(Z_1 = 1) = 1/2$.

(b) Since $h_2(p) = -p\log p - (1-p)\log(1-p) = H(X_i)$,

$$
\begin{aligned}
nh_2(p) &= nH(X_i) && (29)\\
&= H(X_1,\dots,X_n) \text{ [Independence of } X_i] && (30)\\
&\geq H(f(X_1,\dots,X_n)) \text{ [Data Processing Inequality]} && (31)\\
&= H(Z_1,\dots,Z_K,K) && (32)\\
&= H(K) + H(Z_1,\dots,Z_K|K) && (33)\\
&= H(K) + \sum_k p(K=k)H(Z_1,\dots,Z_K|K=k) && (34)\\
&= H(K) + \sum_k p(K=k)k \text{ [} Z_1,\dots,Z_k \text{ are i.i.d and fair when } K=k] && (35)\\
&= H(K) + E[K] && (36)\\
&\geq E[K] && (37)
\end{aligned}
$$

(c) when $n = 4$, $(X_1,\dots,X_4)$ have 16 outcomes with probabilities:

$$
\begin{aligned}
&1 \text{ case} : \Pr(0000) && = && (1-p)^4 && (38)\\
&4 \text{ cases} : \Pr(0001) && = \cdots = \Pr(1000) = && p(1-p)^3 && (39)\\
&6 \text{ cases} : \Pr(0011) && = \cdots = \Pr(1100) = && p^2(1-p)^2 && (40)\\
&4 \text{ cases} : \Pr(0111) && = \cdots = \Pr(1110) = && p^3(1-p) && (41)\\
&1 \text{ case} : \Pr(1111) && = && p^4 && (42)
\end{aligned}
$$

Now we can define the function as follows to get i.i.d. bits and produce as many bits we can:

$$
\begin{aligned}
f(0000) = f(1111) &= \lambda && (43)\\
f(0011) &= 1 && (44)\\
f(1100) &= 0 && (45)\\
f(1001) = f(1110) = f(0001) &= 00 && (46)\\
f(1010) = f(1101) = f(0010) &= 01 && (47)\\
f(0110) = f(1011) = f(0100) &= 10 && (48)\\
f(0101) = f(0111) = f(1000) &= 11 && (49)
\end{aligned}
$$

Problem 5: Extremal characterization for Rényi entropy

Given $s \geq 0$, and a random variable U taking values in $\mathcal{U}$, with probabilitis $p(u)$, consider the distribution $p_s(u) = p(u)^s/Z(s)$ with $Z(s) = \sum_u p(u)^s$.

(a) Show that for any distribution q on $\mathcal{U}$,

$$(1-s)H(q) - sD(q\|p) = -D(q\|p_s) + \log Z(s).$$

(b) Given s and p, conclude that the left hand side above is maximized by the choice by $q = p_s$ with the value $\log Z(s)$,

The quantity

$$H_s(p) := \frac{1}{1-s}\log Z(s) = \frac{1}{1-s}\log\sum_u p(u)^s$$

is known as the *Rényi entropy of order s of the random variable U*. When convenient, we will also write $H_s(U)$ instead of $H_s(p)$.

(c) Show that if U and V are independent random variables

$$H_s(UV) := H_s(U) + H_s(V).$$

[Here UV denotes the pair formed by the two random variables — not their product. E.g., if $\mathcal{U} = \{0,1\}$ and $\mathcal{V} = \{a,b\}$, UV takes values in $\{0a, 0b, 1a, 1b\}$.]

Solution

(a) We start from the left hand side of the equation:

$$\begin{aligned}
(1-s)H(q) - sD(q\|p) &= (1-s)\sum_u q(u)\log\frac{1}{q(u)} - s\sum_u q(u)\log\frac{q(u)}{p(u)} && (50)\\
&= \sum_u q(u)\left((1-s)\log\frac{1}{q(u)} - s\log\frac{q(u)}{p(u)}\right) && (51)\\
&= \sum_u q(u)\log\frac{p(u)^s}{q(u)} && (52)\\
&= \sum_u q(u)\log\frac{p_s(u)Z(s)}{q(u)} && (53)\\
&= \sum_u q(u)\log\frac{p_s(u)}{q(u)} + \sum_u q(u)\log Z(s) && (54)\\
&= -D(q\|p_s) + \log Z(s) && (55)
\end{aligned}$$

(b) We know that $D(q\|p_s) \geq 0$, where equality achieves for $q = p_s$. The left hand side of above equation is maximized when $q = p_s$ and has value $\log Z(s)$.

(c) Since U and V are independent random variables, we have $p(u,v) = p(u)p(v)$.

$$\begin{aligned}
H_s(UV) &= \frac{1}{1-s}\log\sum_{u,v} p(u,v)^s && (56)\\
&= \frac{1}{1-s}\log(\sum_u p(u)^s \sum_v p(v)^s) && (57)\\
&= \frac{1}{1-s}\log\sum_u p(u)^s + \frac{1}{1-s}\log\sum_v p(v)^s && (58)\\
&= H_s(U) + H_s(V) && (59)
\end{aligned}$$

Problem 6: Guessing and Rényi entropy

Suppose X is a random variable taking K values $\{a_1, \dots, a_K\}$ with $p_i = \Pr\{X = a_i\}$. We wish to guess X by asking a sequence of binary questions of the type 'Is $X = a_i$?' until we are answered 'yes'. (Think of guessing a password).

A *guessing strategy* is an ordering of the K possible values of X; we first ask if X is the first value; then if it is the second value, etc. Thus the strategy is described by a function $G(x) \in \{1, \dots, K\}$ that gives the position (first, second, ... Kth) of x in the ordering. I.e., when $X = x$, we ask $G(x)$ questions to guess the value of X. Call G the guessing function of the strategy.

For the rest of the problem suppose $p_1 \geq p_2 \geq \cdots \geq p_K$.

(a) Show that for any guessing function G, the probability of asking fewer than i questions satisfies

$$\Pr(G(X) \le i) \le \sum_{j=1}^{i} p_j$$

and equality holds for the guessing function G^* with $G^*(a_i) = i$, $i = 1, \dots, K$; this is the strategy that first guesses the most probable value a_1, then the next most probable value a_2, etc.

(b) Show that for any increasing function $f : \{1, \dots, K\} \to \mathbb{R}$, $E[f(G(X))]$ is minimized by choosing $G = G^*$. [Hint: $E[f(G(X))] = \sum_{i=1}^{K} f(i) \Pr(G = i)$. Write $\Pr(G = i) = \Pr(G \le i) - \Pr(G \le i-1)$, to write the expectation in terms of $\sum_i [f(i) - f(i+1)] \Pr(G \le i)$, and use (a).]

(c) For any i and $s \ge 0$ prove the inequalities

$$i \le \sum_{j=1}^{i} (p_j/p_i)^s \le \sum_j (p_j/p_i)^s$$

(d) For any $\rho \ge 0$, show that

$$E[G^*(X)^\rho] \le \left(\sum_i p_i^{1-s\rho}\right)\left(\sum_j p_j^s\right)^\rho.$$

for any $s \ge 0$. [Hint: write $E[G^*(X)^\rho] = \sum_i p_i i^\rho$, and use (c) to upper bound i^ρ]

(e) By a choosing s carefully, show that

$$E[G^*(X)^\rho] \le \left(\sum_i p_i^{1/(1+\rho)}\right)^{1+\rho} = \exp\bigl[\rho H_{1/(1+\rho)}(X)\bigr].$$

(f) Suppose $U_1, \dots, U_n$ are i.i.d., each with distribution p, and $X = (U_1, \dots, U_n)$. (I.e., we are trying to guess a password that is made of n independently chosen letters.) Show that

$$\frac{1}{n\rho} \log E[G^*(U_1, \dots, U_n)^\rho] \le H_{1/(1+\rho)}(U_1)$$

[Hint: first observe that $H_\alpha(X) = nH_\alpha(U_1)$. In other words, the ρ-th moment of the number of guesses grows exponentially in n with a rate upper bounded by in terms of the Rényi entropy of the letters.

It is possible a lower bound to $E[G^(U_1, \dots, U_n)^\rho]$ that establishes that the exponential upper bound we found here is asympototically tight.

Solution

(a) The event that $G(X) \le i$ contains the probability of i distinct values.

$$\Pr(G(X) \le i) = \sum_{j=1}^{i} \Pr(G(X) = j) \le \sum_{j=1}^{i} p_j \tag{60}$$

as $p_1, \dots, p_i$ are the i largest probabilities. Equality holds for G^*, since $\Pr(G^* = i) = p_i$.

(b) Note that $\Pr(G(X) \le 0) = 0$ and $\Pr(G(X) \le K) = 1$.

$$\begin{aligned} E[f(G(X))] &= \sum_{i=1}^{K} \Pr(G(X) = i) f(i) && (61) \\ &= \sum_{i=1}^{K} (\Pr(G(X) \le i) - \Pr(G(X) \le i-1)) f(i) && (62) \\ &= \sum_{i=1}^{K-1} \Pr(G(X) \le i)(f(i) - f(i+1)) + f(K) && (63) \\ &\ge \sum_{i=1}^{K-1} \sum_{j=1}^{i} p_j (f(i) - f(i+1)) + f(K) && (64) \end{aligned}$$

where each $\Pr(G(X) \le i) \le \sum_{j=1}^{i} p_j$ with equality holding for $G = G^*$ according to (a) and $f(i) - f(i+1)) \le 0$ since f is an increasing function. Hence, $E[f(G(X))]$ is minimized when $G = G^*$.

(c) Suppose we a distribution with probabilities $\{p_1, \ldots, p_K\}$. For any $i \in \{1, \ldots, K\}$ and $s > 0$:

$$i = \sum_{j=1}^{i} 1^s \le \sum_{j=1}^{i} (p_j/p_i)^s \le \sum_{j=1}^{i} (p_j/p_i)^s + \sum_{j=i+1}^{K} (p_j/p_i)^s = \sum_{j} (p_j/p_i)^s \qquad (65)$$

where the first inequality holds because $p_j/p_i \ge 1$ for each $1 \le j \le i$.

(d)

$$E[G^*(X)^\rho] = \sum_i \Pr(G^*(X) = i) i^\rho = \sum_i p_i i^\rho \le \sum_i p_i \left(\sum_j \frac{p_j^s}{p_i^s} \right)^\rho = \left(\sum_i p_i^{1-s\rho} \right) \left(\sum_j p_j^s \right)^\rho \qquad (66)$$

(e) Since inequality (66) holds for any $s > 0$, we can choose $s = \frac{1}{1+\rho}$ and get

$$\begin{aligned} E[G^*(X)^\rho] &\le \left(\sum_i p_i^{\frac{1}{1+\rho}} \right) \left(\sum_j p_j^{\frac{1}{1+\rho}} \right)^\rho && (67) \\ &= \left(\sum_i p_i^{\frac{1}{1+\rho}} \right)^{1+\rho} && (68) \\ &= \exp \left[(1+\rho) \log \sum_i p_i^{\frac{1}{1+\rho}} \right] && (69) \\ &= \exp \left[\rho \frac{1}{1 - \frac{1}{1+\rho}} \log \sum_i p_i^{\frac{1}{1+\rho}} \right] && (70) \\ &= \exp \left[\rho H_{1/(1+\rho)}(X) \right] && (71) \end{aligned}$$

(f) Follow the hint that $H_\alpha(X) = n H_\alpha(U_1)$:

$$\begin{aligned} \frac{1}{n\rho} \log E[G^*(U_1, \ldots, U_n)^\rho] &\le \frac{1}{n\rho} \log \exp[\rho H_{1/(1+\rho)}(X)] && (72) \\ &= \frac{1}{n} H_{1/(1+\rho)}(X) && (73) \\ &= H_{1/(1+\rho)}(U_1) && (74) \end{aligned}$$