Tarea No. 1: Solución

1. Definamos $f(x) = \ln x - (x-1)$. Debemos demostrar que $f(x)$ tiene un máximo y que ese valor máximo es 0.

   En efecto,

$$f'(x) = \frac{1}{x} - 1 \Rightarrow x = 1 \text{ es un punto óptimo.}$$
$$f''(x) = -\frac{1}{x^2} < 0 \; \forall \, x.$$

   Por lo tanto $f(1)$ es un máximo:

$$f(x) \leq f(1) = \ln 1 - (1-1) = 0,$$

   es decir, $\ln x \leq x - 1$.

   Para la segunda parte consideramos $y = 1/x$ y aplicamos la desigualdad que acabamos de demostrar:

$$\begin{aligned} \ln\left(\frac{1}{x}\right) &\leq \frac{1}{x} - 1 \\ -\ln x &\leq \frac{1}{x} - 1 \\ \ln x &\geq 1 - \frac{1}{x}. \end{aligned}$$

2. Sea $P_\lambda[k] = e^{-\lambda}\dfrac{\lambda^k}{k!}$.

$$\begin{aligned} H(P_\lambda) &= -\sum_{k=0}^{\infty} P_\lambda[k] \log P_\lambda[k] \\ &= -\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} \log\left(e^{-\lambda}\frac{\lambda^k}{k!}\right) \\ &= -e^{-\lambda}\left[\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log(e^{-\lambda}) + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log\left(\frac{\lambda^k}{k!}\right)\right] \\ &= -e^{-\lambda}\left[\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}(-\lambda \log e + k \log \lambda) - \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log k!\right] \\ &= -e^{-\lambda}\left[-\lambda \log e \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \log \lambda) \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}\right] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log k! \\ &= -e^{-\lambda}\Big[-\lambda \log(e) \cdot e^{\lambda} + \lambda \log \lambda \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{e^\lambda} \Big] + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log k! \end{aligned}$$

$$\boxed{H(P_\lambda) = \lambda \log \frac{e}{\lambda} + e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \log k!}$$

3. Recordamos que $H(p) = -\sum_j p_j \log p_j = \mathbb{E}[-\log p]$. Por lo tanto,

$$\begin{aligned}
H(\mathbf{p}^n) &= \mathbb{E}[-\log \mathbf{p}^n] \\
&= \mathbb{E}[-\log \prod_{k=1}^{n} p(a_{jk})] \\
&= \mathbb{E}[-\sum_{k=1}^{n} \log p(a_{jk})] \\
&= \sum_{k=1}^{n} \mathbb{E}[-\log p(a_{jk})] \\
&= \sum_{k=1}^{n} H(p) \\
&= nH(p).
\end{aligned}$$

4. Si definimos $Y = g(X)$, Y es una variable aleatoria si y sólo si $g(\cdot)$ es medible. Asumiendo esta hipótesis, tenemos

$$\begin{aligned}
H(X,Y) &= H(X) + H(Y|X) \\
\Rightarrow H(X, g(X)) &= H(X) + H(g(X)|X).
\end{aligned}$$

Notemos que si X es conocido, la incertidumbre acerca de $g(X)$ es cero. Esto es, $H(g(X)|X) = 0$. Por lo tanto,

$$H(X, g(X)) = H(X).$$

Por otro lado,

$$\begin{aligned}
H(X,Y) = H(Y,X) &= H(Y) + H(X|Y) \\
\Rightarrow H(X, g(X)) &= H(g(X)) + H(X|g(X)).
\end{aligned}$$

Pero la entropía condicional $H(X|g(X)) \geq 0$. Notemos que si $g(X)$ es una medición de X, puede que no resuelva toda la incertidumbre acerca de X. Por ello

$$H(X) = H(X, g(X)) = H(g(X)) + H(X|g(X)) \geq H(g(X)).$$

Lo que nos permite concluir $H(X) \geq H(g(X))$.

5. Queremos resolver el problema

$$\max_{\substack{\text{sujeto a:} \\ \sum_j jp_j = A \\ \sum_j p_j = 1}} \; -\sum_{j=0}^{\infty} p_j \log p_j \tag{1}$$

La primera restricción corresponde al valor esperado de la función masa de probabilidad, y la segunda corresponde a la condición de normalización que debe cumplir $\mathbf{p}$. En lo que sigue asumiremos $A > 0$.

Definimos el lagrangiano

$$\mathcal{L}(\mathbf{p}, \lambda, \mu) = -\sum_{j=0}^{\infty} p_j \log p_j - \lambda\Big(\sum_{j=0}^{\infty} jp_j - A\Big) - \mu\Big(\sum_{j=0}^{\infty} p_j - 1\Big).$$

Calculamos la $\partial\mathcal{L}/\partial p_k$:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p_k} &= \frac{\partial}{\partial p_k}\Big[p_k \log p_k - \lambda k p_k - \mu p_k\Big] \\
&= -\log p_k - \frac{p_k}{p_k} + \lambda k + \mu.
\end{aligned}$$

Igualamos a 0 para obtener una condición para el óptimo:

$$-\log p_k - 1 + \lambda k + \mu = 0$$
$$\Rightarrow p_j = e^{\lambda j} e^{\mu - 1}.$$

Para poder eliminar λ y μ debemos aplicar las condiciones de Kuhn-Tucker. Ello nos lleva a las siguientes ecuaciones:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \sum_{j=0}^{\infty} p_j - 1 = 0,$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{j=0}^{\infty} j p_j - A = 0.$$

De la primera ecuación podemos despejar μ.

$$\sum_{j=0}^{\infty} p_j = \sum_{j=0}^{\infty} e^{\lambda j} e^{\mu-1} = 1$$
$$\Rightarrow e^{\mu-1} = \Big(\sum_{j=0}^{\infty} e^{\lambda j}\Big)^{-1}$$
$$\Rightarrow p_j = e^{\lambda j}\Big(\sum_{j=0}^{\infty} e^{\lambda j}\Big)^{-1}.$$

Notemos que si $e^{\lambda} < 1$, la serie converge a

$$\sum_{j=0}^{\infty} e^{\lambda j} = \sum_{j=0}^{\infty} \left(e^{\lambda}\right)^{j} = \frac{1}{1 - e^{\lambda}}.$$

Por lo tanto,

$$p_j = (1 - e^{\lambda}) e^{\lambda j}.$$

Nos queda la segunda ecuación:

$$\sum_{j=0}^{\infty} j p_j = A.$$

Recordamos que $\sum_{k=0}^{\infty} k r^k = r \frac{d}{dr} \sum_{k=0}^{\infty} r^k = \frac{r}{(1-r)^2}$, por lo que

$$\begin{aligned}
\sum_{j=0}^{\infty} j p_j &= \sum_{j=0}^{\infty} j (1 - e^{\lambda}) e^{\lambda j} \\
&= (1 - e^{\lambda}) \sum_{j=0}^{\infty} j e^{\lambda j} \\
&= (1 - e^{\lambda}) \frac{e^{\lambda}}{(1 - e^{\lambda})^2} \\
&= \frac{e^{\lambda}}{1 - e^{\lambda}} \\
&= A \\
\Rightarrow e^{\lambda} &= \frac{A}{1 + A}.
\end{aligned}$$

Despejando obtenemos

$$p_j = \Big(\frac{A}{1+A}\Big)^{j} \Big(1 - \frac{A}{1+A}\Big) = \Big(\frac{A}{1+A}\Big)^{j} \frac{1}{1+A}.$$

Notemos que si identificamos $q = \dfrac{A}{1+A}$, entonces $p_j = q^j(1-q)$ corresponde a una distribución geométrica de parámetro q.
La entropía máxima se alcanza porque $H(p)$ es cóncava en p. En este caso,

$$\begin{aligned}
H(p) &= -\sum_{j=0}^{\infty} q^j(1-q)\log(q^j(1-q)) \\
&= -(1-q)\sum_{j=0}^{\infty} q^j[\log q^j - \log(1-q)] \\
&= -(1-q)\log q \sum_{j=0}^{\infty} jq^j + \log(1-q)\sum_{j=0}^{\infty} q^j \\
&= -(1-q)\frac{q}{(1-q)^2}\log q + \log(1-q)\frac{1}{1-q} \\
\Rightarrow H_{\text{max}} &= -\frac{q}{1-q}\log q + \frac{1}{1-q}\log(1-q).
\end{aligned}$$