Modern Methods in Applied Statistics
Homework 3

Daniel F. Noriega

April 15th, 2024

1 Exponential-gamma conjugacy

1.1 Exponential family forms

Recall general exponential family distribution form:

$$P(x|\eta) = h(x)\exp(\eta^T t(x) - \alpha(\eta)) \tag{1}$$

Exponential

We want an exponential family distribution of the form $P(x|\mu) = \mu\ exp(-\mu)$. By setting the following into (1) above:

$$h(x) = 1 \text{ for } x > 0;\ t(x) = x \text{ and } \eta = -\mu$$

$$\text{we get: } P(x|\eta) = 1\exp(-\mu x - \alpha(\eta))$$

$$\text{Setting } \alpha(\eta) = -log(-\eta)$$

$$\begin{aligned} P(x|\mu) &= \exp(-\mu x + log(\mu)) \\ &= \mu\exp(-\mu x) \text{ [functional form desired]} \end{aligned}$$

Gamma

We want an exponential family distribution of the form

$$P(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} exp(-bx)$$

By setting the following into (1) above:

$$h(x) = 1 \text{ for } x > 0;\ t(x) = (-x, log(x)) \text{ and } \eta = (b, a-1)$$

$$\text{we get: } P(x|\eta) = 1\exp(-bx + (a-1)log(x) - \alpha(b, a-1))$$

$$\text{Setting } \alpha(\eta) = -a.log(b) + log(\Gamma(a))$$

$$\begin{aligned} P(x|\eta) &= x^{a-1}\exp[-bx + a.log(b) - log(\Gamma(a))] \\ &= \frac{b^a}{\Gamma(a)} x^{a-1}\exp(-bx) \text{ [functional form desired]} \end{aligned}$$

1.2 Conjugacy

Using a conjugate prior for η taking the form

$$P(\eta|\lambda) = h_c(\eta)\exp(\lambda_1^T\eta - \lambda_2 a_\ell(\eta) - a_c(\lambda)) \tag{2}$$

where $\lambda = [\lambda_1, \lambda_2]$ and $t(\eta) = [\eta, -a_\ell(\eta)]$, we want to show that the prior is a Gamma distribution such that:

$$\Gamma^{form}{}_{posterior} = \Gamma^{form}{}_{prior}.\mu\exp(-\mu x)$$

Given the prior and likelihood, the posterior becomes:

$$\begin{aligned} posterior &= \frac{b^a}{\Gamma(a)}\mu^{a-1}\exp(-b\mu).\mu\exp(-\mu x) \\ &= \frac{b^a}{\Gamma(a)}\mu^{a}\exp(-\mu(b+x)) \end{aligned}$$

Note that the functional Gamma form used for the prior is similar to equation (2):

$$\frac{b^a}{\Gamma(a)}\mu^{a-1}\exp(-b\mu) \equiv h_c(\eta)\exp(\lambda_1^T\eta - \lambda_2 a_\ell(\eta) - a_c(\lambda))$$

If we set:

$$\eta \triangleq \mu, \quad h_c(\eta) \triangleq 1, \quad a_c(\lambda) \triangleq -a.log(b) + log(\Gamma(a))$$

$$\lambda_1^T\eta \triangleq -b\mu, \quad \lambda_2 a_\ell(\eta) \triangleq -(a-1).log(\mu)$$

And the posterior would be a Gamma functional form:

$$posterior = \frac{b^a}{\Gamma(a)}\mu^{a}\exp(-\mu(b+x)) \equiv \frac{b'^{a'}}{\Gamma(a')}x^{a'-1}\exp(-b'x)$$

By updating:

Given observed x, $\quad a' = a+1, \quad b' = b+x$

This process adjusts the parameters a' and b' based on the new evidence x. If the likelihood $P(x|\mu)$ is exponential, then the conjugate prior for μ is a Gamma distribution, resulting in a Gamma-distributed posterior.

1.3 Posterior

Given n *i.i.d.* observations from an exponential distribution, the **likelihood** of observing $x_1, x_2, \ldots, x_n$ given μ is:

$$L(x_{1:n}|\mu) = \prod_{i=1}^{n}\mu\exp(-\mu x_i) = \mu^n\exp(-\mu\sum_{i=1}^{n}x_i)$$

The **prior** distribution for μ is a gamma distribution with parameters a and b, given by:

$$P(\mu|a,b) = \frac{b^a}{\Gamma(a)}\mu^{a-1}\exp(-b\mu)$$

The **posterior** distribution $P(\mu|x_{1:n},a,b)$ would be given by (unnormalized):

$$\begin{aligned}
P(\mu|x_{1:n},a,b) &\propto L(x_{1:n}|\mu) \times P(\mu|a,b) \\
&\propto \mu^n \exp(-\mu\sum_{i=1}^{n} x_i) \times \frac{b^a}{\Gamma(a)}\mu^{a-1}\exp(-b\mu) \\
&\propto \mu^{n+a-1}\exp\left(-\mu(b+\sum_{i=1}^{n} x_i)\right)
\end{aligned}$$

Which is in the form of a gamma distribution $Gamma(\mu; a', b')$ where the updated (posterior) parameters a' and b' are given by:

$$\begin{aligned}
a' &= a+n \\
b' &= b+\sum_{i=1}^{n} x_i
\end{aligned}$$

1.4 Prior predictive distribution

To derive $P(x_1|a,b) = \int P(x_1|\mu)P(\mu|a,b)d\mu$, we have:

$$\begin{aligned}
P(x_1|a,b) &= \int_0^\infty \mu\exp(-\mu x_1)\frac{b^a}{\Gamma(a)}\mu^{a-1}\exp(-b\mu)d\mu \\
&= \frac{b^a}{\Gamma(a)}\int_0^\infty \mu^a\exp(-\mu(b+x_1))d\mu
\end{aligned}$$

We recognize the integral as the integral function of a gamma distribution (which is of the form $\int_0^\infty t^k e^{-st}dt = \frac{\Gamma(k+1)}{s^{k+1}}$), so that:

$$\begin{aligned}
P(x_1|a,b) &= \frac{b^a}{\Gamma(a)}\int_0^\infty \mu^a\exp(-\mu(b+x_1))d\mu \\
&= \frac{b^a}{\Gamma(a)}\frac{\Gamma(a+1)}{(b+x_1)^{a+1}} = \frac{b^a}{\Gamma(a)}\frac{a\Gamma(a)}{(b+x_1)^{a+1}} \\
&= \frac{b^a a}{(b+x_1)^{a+1}}
\end{aligned}$$

1.5 Posterior predictive distribution

To derive $P(x_{n+1}|x_{1:n}a, b) = \int P(x_{n+1}|\mu)P(\mu|x_{1:n}a, b)d\mu$, we have:

$$\begin{aligned} P(x_{n+1}|x_{1:n}, a, b) &\propto \int_0^\infty \mu \exp(-\mu x_{n+1})\mu^{n+a-1} \exp\left(-\mu(b + \sum_{i=1}^{n} x_i)\right) \\ &\propto \int_0^\infty \mu^{n+a} \exp\left(-\mu(b + \sum_{i=1}^{n+1} x_i)\right) \end{aligned}$$

Which we can again recognize as the integral function of a gamma distribution (of the form $\int_0^\infty t^k e^{-st} dt = \frac{\Gamma(k+1)}{s^{k+1}}$), so that:

$$\begin{aligned} P(x_{n+1}|x_{1:n}, a, b) &= \frac{(b + \sum_{i=1}^{n} x_i)^{a+n}}{\Gamma(a+n)} \int_0^\infty \mu^{n+a} \exp\left(-\mu(b + \sum_{i=1}^{n+1} x_i)\right) \\ &= \frac{(b + \sum_{i=1}^{n} x_i)^{a+n}}{\Gamma(a+n)} \frac{\Gamma(n+a+1)}{(b + \sum_{i=1}^{n+1} x_i)^{n+a+1}} \\ &= \frac{(n+a)(b + \sum_{i=1}^{n} x_i)^{a+n}}{(b + \sum_{i=1}^{n+1} x_i)^{n+a+1}} \end{aligned}$$

2 Gamma-Poisson and entropy

2.1 Posterior

To derive the posterior $P(\mu|x,a,b)$ we have:

Given $x \sim Pois(\mu), \quad \mu \sim Gam(a,b)$:

$$\begin{aligned}
P(\mu|x,a,b) &\propto P(x|\mu)P(\mu|a,b) \\
&\propto \frac{\mu^x}{x!}\exp(-\mu)\frac{b^a}{\Gamma(a)}\mu^{a-1}\exp(-b\mu) \\
&\propto \frac{b^a \mu^{x+a-1}}{x!\Gamma(a)}\exp(-\mu(b+1)) \\
&\propto_\mu \mu^{x+a-1}\exp(-\mu(b+1))
\end{aligned}$$

Which we recognize as a *Gamma* distribution with parameters $a' = a + x$ and $b' = b + 1$, so that:

$$P(\mu|x,a,b) \sim Gam(a+x, b+1)$$

2.2 Exponential family forms

Including equation (1) below for reference:

$$P(x|\eta) = h(x)\exp(\eta^T t(x) - \alpha(\eta)) \qquad (1)$$

Poisson

We want an exponential distribution of the form $P(x|\mu) = \frac{\mu^x}{x!}\exp(-\mu)$. By setting the following into (1) above:

$$h(x) = \frac{1}{x!},\ x \in \mathbb{N}_0;\ t(x) = x \text{ and } \eta = log(\mu)$$

$$\text{we get:} \quad P(x|\mu) = \frac{1}{x!}\exp(log(\mu)x - \alpha(log(\mu)))$$

$$\text{Setting} \quad \alpha(\eta) = e^\eta = \mu$$

$$\begin{aligned}
P(x|\mu) &= \frac{1}{x!}\exp(log(\mu)x - \mu) \\
&= \frac{\mu^x}{x!}\exp(-\mu) \text{ [functional form desired]}
\end{aligned}$$

Negative Binomial Distribution (with known r)

We want an exponential distribution of the form $P(x|p) = \frac{\Gamma(x+r)}{x!\Gamma(r)}(1-p)^r p^x$, with $r > 0$ being an arbitrary constant.

We can re-write $P(x|p)$ as:

$$P(x|p) = \frac{\Gamma(x+r)}{x!\Gamma(r)}(1-p)^r p^x = \frac{\Gamma(x+r)}{x!\Gamma(r)} \exp\left(r\ log(1-p) + x\ log(p)\right)$$

By setting the following into (1) above:

$$h(x) = \frac{\Gamma(x+r)}{x!\Gamma(r)},\ x \in \mathbb{N}_0; \quad t(x) = x; \quad \eta = log(p)$$
$$\text{and} \quad \alpha(\eta) = -r\ log(1-\exp(\eta))$$

$$\begin{aligned} P(x|p) &= \frac{\Gamma(x+r)}{x!\Gamma(r)} \exp(log(p)x + r\ log[1-\exp(log(p))]) \\ &= \frac{\Gamma(x+r)}{x!\Gamma(r)} \exp(log(p)x + r\ log[1-p]) \\ &= \frac{\Gamma(x+r)}{x!\Gamma(r)} p^x (1-p)^r \text{ [functional form desired]} \end{aligned}$$

2.3 KL divergence between two Poissons

Given $KL(P(x) \parallel Q(x)) = \sum_{x\in\mathcal{X}} P(x)\ log\left[\frac{P(x)}{Q(x)}\right]$, the KL divergence between two Poissons would be given by:

$$\begin{aligned} KL(Pois(x;\mu_1) \parallel Pois(x;\mu_2)) &= \sum_{x\in\mathcal{X}} Pois(x;\mu_1) log\left[\frac{Pois(x;\mu_1)}{Pois(x;\mu_2)}\right] \\ &= \sum_{x\in\mathcal{X}} \frac{\mu_1^x}{x!}\exp(-\mu_1) log\left[\frac{\frac{\mu_1^x}{x!}\exp(-\mu_1)}{\frac{\mu_2^x}{x!}\exp(-\mu_2)}\right] \\ &= \sum_{x\in\mathcal{X}} \frac{\mu_1^x}{x!}\exp(-\mu_1) log\left[(\frac{\mu_1}{\mu_2})^x \exp(\mu_2-\mu_1)\right] \\ &= \sum_{x\in\mathcal{X}} \frac{\mu_1^x}{x!}\exp(-\mu_1)\left[x\ log\left(\frac{\mu_1}{\mu_2}\right) + \mu_2 - \mu_1\right] \end{aligned}$$

Considering that $x \in \mathbb{N}_0$:

$$= \sum_{x=0}^{\infty} \frac{\mu_1^x}{x!}\exp(-\mu_1)\left[x\ log\left(\frac{\mu_1}{\mu_2}\right) + \mu_2 - \mu_1\right]$$

Given that it is a probability distribution: $\sum_{x=0}^{\infty} \frac{\mu_1^x}{x!}\exp(-\mu_1) = 1$, so:

$$\begin{aligned} &= \mu_2 - \mu_1 + log\left(\frac{\mu_1}{\mu_2}\right) \sum_{x=0}^{\infty} \frac{\mu_1^x}{x!}\exp(-\mu_1)\ x \\ &= \mu_2 - \mu_1 + log\left(\frac{\mu_1}{\mu_2}\right)\mu_1 \end{aligned}$$

Where in the last step the sum is just equal to an expectation $\sum_x x \cdot P(x)$, and $\sum_{x=0}^{\infty} \frac{\mu_1^x}{x!} \exp(-\mu_1)\ x = \mu_1$. So,

$$KL(Pois(x;\mu_1) \parallel Pois(x;\mu_2)) = \mu_2 - \mu_1 + log\left(\frac{\mu_1}{\mu_2}\right)\mu_1$$

2.4 Poisson and negative binomial entropies

Defining and simplifying intermediate expressions:

2.4.1 $H\left(\mathbf{NB}\left(a, \frac{1}{1+b}\right)\right)$:

$$= -\sum_{x=0}^{\infty} \mathrm{NB}\left(x; a, \frac{1}{1+b}\right) log\left[\mathrm{NB}(x:a, \frac{1}{1+b})\right]$$

$$= -\sum_{x=0}^{\infty} \frac{\Gamma(x+a)}{x!\Gamma(a)}\left(\frac{b}{1+b}\right)^a \left(\frac{1}{1+b}\right)^x log\left[\frac{\Gamma(x+a)}{x!\Gamma(a)}\left(\frac{b}{1+b}\right)^a \left(\frac{1}{1+b}\right)^x\right]$$

Given $\sum_{x=0}^{\infty} \mathrm{NB}\left(x; a, \frac{1}{1+b}\right) = 1$ and $\mathbb{E}\left[\mathrm{NB}\left(x; a, \frac{1}{1+b}\right)\right] = ab$:

$$= -log\left(\frac{b}{1+b}\right)a - log\left(\frac{1}{1+b}\right)ab - \sum_{x=0}^{\infty} \mathrm{NB}\left(x; a, \frac{1}{1+b}\right) log\left[\frac{\Gamma(x+a)}{x!\Gamma(a)}\right]$$

$$= -log\left(\frac{b}{1+b}\right)a - log\left(\frac{1}{1+b}\right)ab$$

$$-\sum_{x=0}^{\infty} \mathrm{NB}\left(x; a, \frac{1}{1+b}\right)\left[log(\Gamma(x+a)) - log(\Gamma(x+1)) - log(\Gamma(a))\right]$$

$$= -a\log\left(\frac{b}{1+b}\right) - ab\log\left(\frac{1}{1+b}\right) - \log(\Gamma(a))$$

$$+\mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] - \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+1))]$$

2.4.2 $H\left(Pois(y)\right)$:

$$= -\sum_{x=0}^{\infty} Pois(x;y) log[Pois(x;y)]$$

$$= -\sum_{x=0}^{\infty} \frac{y^x}{x!}\exp(-y) log\left(\frac{y^x}{x!}\exp(-y)\right)$$

$$= -\sum_{x=0}^{\infty} \frac{y^x}{x!}\exp(-y)\left[x.log(y) - log(x!) - y\right]$$

Given $\sum_{x=0}^{\infty} \frac{y^x}{x!} \exp(-y) = 1$ and $\sum_{x=0}^{\infty} \frac{y^x}{x!} \exp(-y)x = y$:

$$
\begin{aligned}
&= y - log(y)y + \sum_{x=0}^{\infty} \frac{y^x}{x!} \exp(-y) \left[log(x!)\right] \\
&= y - log(y)y + \sum_{x=0}^{\infty} \frac{y^x}{x!} \exp(-y) \left[\sum_{i=1}^{x} log(i)\right] \\
&= y - log(y)y + \sum_{x=0}^{\infty} \frac{y^x}{x!} \exp(-y) \left[log(\Gamma(x+1))\right] \\
&= y - log(y)y + \sum_{x=0}^{\infty} Pois(x;y) \left[log(\Gamma(x+1))\right]
\end{aligned}
$$

2.4.3 $H\left(Pois(y)|Gam(a,b)\right)$:

Given:

$$
\begin{aligned}
H(Pois(y)|Gam(a,b)) &= \int_0^{\infty} H(Pois(y)) \times Gam(y;a,b)\, dy \\
H(Pois(y)) &= -\sum_{x=0}^{\infty} Pois(x;y) \ln(Pois(x;y)) \\
Pois(x;y) &= \frac{y^x e^{-y}}{x!} \\
Gam(y;a,b) &= \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}
\end{aligned}
$$

Substitute and integrate using the Gamma-Poisson mixture property:

$$
\begin{aligned}
H(Pois(y)|Gam(a,b)) &= -\int_0^{\infty} \left(\sum_{x=0}^{\infty} \frac{y^x e^{-y}}{x!} log\left(\frac{y^x e^{-y}}{x!}\right)\right) \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}\, dy \\
&= -\sum_{x=0}^{\infty} \int_0^{\infty} \frac{y^x e^{-y}}{x!} log\left(\frac{y^x e^{-y}}{x!}\right) \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}\, dy \\
&= -\sum_{x=0}^{\infty} \mathrm{NB}(x;a,\frac{1}{1+b}) log(Pois(x;y))
\end{aligned}
$$

$$
\begin{aligned}
&= -\sum_{x=0}^{\infty} \mathrm{NB}(x; a, \frac{1}{1+b}) \left[x \log y - y - \log(x!)\right] \\
&= -\sum_{x=0}^{\infty} \mathrm{NB}(x; a, \frac{1}{1+b}) \left[x \mathbb{E}[\log y] - \mathbb{E}[y] - \log(x!)\right] \\
&= -\sum_{x=0}^{\infty} \mathrm{NB}(x; a, \frac{1}{1+b}) \left[x \log\left(\frac{a}{b}\right) - \frac{a}{b} - \log(x!)\right] \\
&= -\left(\log\left(\frac{a}{b}\right) \mathbb{E}_{\mathrm{NB}}[X] - \frac{a}{b} - \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+1))]\right) \\
&= -\left(\log\left(\frac{a}{b}\right) \frac{a}{b} - \frac{a}{b} - \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+1))]\right) \\
H(Pois(y)|Gam(a,b)) &= -\frac{a}{b}\left(\log\left(\frac{a}{b}\right) - 1\right) - \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+1))]
\end{aligned}
$$

2.4.4 Comparing (establishing bound)

Comparing $H(Pois(y)|Gam(a,b))$ with $H(\mathrm{NB}(a, \frac{1}{1+b}))$ above:

$$
\begin{aligned}
H(\mathrm{NB}(a, \frac{1}{1+b})) = -a \log\left(\frac{b}{1+b}\right) - ab \log\left(\frac{1}{1+b}\right) - \log(\Gamma(a)) \\
+ \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] - \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+1))]
\end{aligned}
$$

Removing the $\mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+1))]$ term from both sides, we want to conclude that:

$$
\begin{aligned}
-a \log\left(\frac{b}{1+b}\right) - ab \log\left(\frac{1}{1+b}\right) - \log(\Gamma(a)) + \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] \\
\geq \\
-\frac{a}{b}\left(\log\left(\frac{a}{b}\right) - 1\right)
\end{aligned}
$$

This is equivalent to:

$$
\begin{aligned}
a \log\left(\frac{1+b}{b}\right) + ab \log(1+b) - \log(\Gamma(a)) + \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] \\
\geq \\
-\frac{a}{b}\left(\log\left(\frac{a}{b}\right) - 1\right)
\end{aligned}
$$

And:

$$
\begin{aligned}
a \log\left(\frac{a-b}{b}\right) + ab^2 \log(1+b) + ab \log\left(\frac{1+b}{b}\right) - b \log\left(\frac{a!}{a}\right) \\
+ b \cdot \mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] \geq 0
\end{aligned}
$$

Given the convexity of $\log(\Gamma(x))$, Jensen's Inequality implies:

$$\mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] \geq \log(\Gamma(\mathbb{E}_{\mathrm{NB}}[X+a]))$$

Since $\mathbb{E}[X] = ab$ for $X \sim \mathrm{NB}(a, \frac{1}{1+b})$, we have $\mathbb{E}_{\mathrm{NB}}[X+a] = ab + a$. Thus:

$$\mathbb{E}_{\mathrm{NB}}[\log(\Gamma(X+a))] \geq \log(\Gamma(ab+a))$$

indicating a positive and significant contribution due to the properties of the Gamma function.

So we want to show that:

$$a \log\left(\frac{a-b}{b}\right) + ab^2 \log(1+b) + ab \log\left(\frac{1+b}{b}\right) - b \log\left(\frac{a!}{a}\right) \\ + b \cdot \log(\Gamma(ab+a))] \geq 0$$

All the terms in the expression above are positive (under $a > b$, otherwise undefined), except for $-b \log\left(\frac{a!}{a}\right)$. However, we can show the other terms will dominate (in fact, I will only use one of them below).

Using Stirling's approximation:

$$\log(n!) \approx n \log n - n + \frac{1}{2} \log(2\pi n)$$

and thus,

$$\log\left(\frac{a!}{a}\right) \approx (a-1) \log a - a + \frac{1}{2} \log(2\pi a)$$

For the gamma function term, we use:

$$\log(\Gamma(ab+a)) \approx (ab+a) \log(ab+a) - (ab+a) + \frac{1}{2} \log(2\pi(ab+a))$$

Comparing the dominant terms:

$$(ab+a) \log(ab+a) > (a-1) \log a$$

so that $\log(\Gamma(ab+a))$ is greater than $\log\left(\frac{a!}{a}\right)$ for $a > b > 0$.

So, we can indeed assert that:

$$a \log\left(\frac{a-b}{b}\right) + ab^2 \log(1+b) + ab \log\left(\frac{1+b}{b}\right) - b \log\left(\frac{a!}{a}\right) \\ + b \cdot \log(\Gamma(ab+a))] \geq 0$$

Which implies that:

$$H(Pois(y)|Gam(a,b)) \geq H(Pois(y)|Gam(a,b)) \quad \textbf{[as desired].}$$

3 Where should we search next?

3.1 Minimizing uncertainty

Selecting a k^* to search in such that $k^* = \arg\max_k H(Z) - H(Z|Y_k)$ would be equivalent to:

$$\begin{aligned}
k^* &= \arg\max_k H(Z) - H(Z|Y_k) \\
&= \arg\max_k H(Z) - [P(Y_k = 0)H(Z|Y_k = 0) + P(Y_k = 1)H(Z|Y_k = 1)] \\
&= \arg\max_k H(Z) - [(1 - \pi_k)H(Z) + \pi_k \cdot 0] \\
&= \arg\max_k \pi_k H(Z)
\end{aligned}$$

So that searching the most likely cell maximizes the expected reduction in uncertainty:

$$k^* = \arg\max_k \pi_k$$

It is worth noting that $H(Z|Y_k = 1) = 0$ since we find the submarine with certainty if searching the right cell.

Additionally, we can see that $H(Z|Y_k = 0) \approx H(Z)$ because when the search in cell k fails, the probabilities for other cells are adjusted:

$$\pi_j' = \frac{\pi_j}{1 - \pi_k}, \quad j \neq k$$

We can recalculate the conditional entropy $H(Z|Y_k = 0)$:

$$\begin{aligned}
H(Z|Y_k = 0) &= -\sum_{j \neq k} \pi_j' \log(\pi_j') \\
&= -\sum_{j \neq k} \frac{\pi_j}{1 - \pi_k} \log\left(\frac{\pi_j}{1 - \pi_k}\right) \\
&= -\sum_{j \neq k} \frac{\pi_j}{1 - \pi_k} \left(\log(\pi_j) - \log(1 - \pi_k)\right) \\
&= \left(-\sum_{j \neq k} \pi_j \log(\pi_j)\right) + \log(1 - \pi_k) \\
&\approx H(Z) + \log(1 - \pi_k)
\end{aligned}$$

If π_k approaches zero (e.g. $K \gg 1$ or it is difficult to narrow down to a few likely cells), the term $\log(1 - \pi_k)$ approaches zero, leading to:

$$H(Z|Y_k = 0) \approx H(Z)$$

Thus, when π_k is small, failing to find the submarine in cell k hardly changes the overall entropy, indicating minimal gain in information.

3.2 Incorporating SEPs

Now assuming $P(Y_k = 1|Z = k) = q$, we would have:

$$
\begin{aligned}
k^* &= \arg\max_k H(Z) - H(Z|Y_k) \\
&= \arg\max_k H(Z) - [P(Y_k = 0)H(Z|Y_k = 0) + P(Y_k = 1)H(Z|Y_k = 1)] \\
&= \arg\max_k H(Z) - [(1 - \pi_k q_k)H(Z|Y_k = 0) + \pi_k q_k H(Z|Y_k = 1)] \\
&= \arg\max_k H(Z) - [(1 - \pi_k q_k)H(Z|Y_k = 0) + \pi_k q_k \cdot 0] \\
&= \arg\max_k H(Z) - [(1 - \pi_k q_k)H(Z)] \\
&= \arg\max_k \pi_k q_k
\end{aligned}
$$

(See simplifications in 3.1 above: $H(Z|Y_k = 1) = 0$; $H(Z|Y_k = 0) \approx H(Z)$)

As a functional form:

$$k^* = \arg\max_k f(\pi_k, q_k) = \arg\max_k \pi_k q_k$$

———-

Alternatively, we could think of minimizing the entropy given a Y_k draw, so that:

$$k^* = \min_k H_2(Y_k) = \min_k H_2(\pi_k q_k)$$

Differentiating $H_2(p) = p\ log_2 \frac{1}{p} + (1-p) log_2 \frac{1}{1-p}$ we see that:

$$
\begin{aligned}
H_2(p) &= -p \log_2(p) - (1-p)\log_2(1-p), \\
\frac{dH_2}{dp} &= -\log_2(p) - 1 - (-\log_2(1-p) - 1) = -\log_2(p) + \log_2(1-p).
\end{aligned}
$$

Setting $\frac{dH_2}{dp} = 0$ to get critical points:

$$
\begin{aligned}
-\log_2(p) + \log_2(1-p) &= 0 \\
\log_2(1-p) &= \log_2(p) \\
1 - p &= p \\
p &= \frac{1}{2}
\end{aligned}
$$

The critical point at $p = \frac{1}{2}$ corresponds to the maximum entropy (the function is concave). Since minimizing $H_2(\pi_k q_k)$ is equivalent to avoiding this point of maximum uncertainty, we aim for values of $\pi_k q_k$ approaching 0 or 1. The practical choice for this search optimization setting is maximizing $\pi_k q_k$ towards 1, hence:

$$k^* = \arg\max_k \pi_k q_k$$

3.3 Example involving SEPs

Using the derived formula $k^* = \arg\max_k \pi_k q_k$, the values for each cell are:

$$\begin{aligned}
&\text{Cell 1:} \quad \pi_1 = \frac{3}{4}, \quad q_1 = \frac{1}{4}, \quad \text{hence} \quad \pi_1 q_1 = \frac{3}{16} \\
&\text{Cell 2:} \quad \pi_2 = \frac{1}{8}, \quad q_2 = \frac{1}{2}, \quad \text{hence} \quad \pi_2 q_2 = \frac{1}{16} \\
&\text{Cell 3:} \quad \pi_3 = \frac{1}{16}, \quad q_3 = \frac{3}{4}, \quad \text{hence} \quad \pi_3 q_3 = \frac{3}{64} \\
&\text{Cell 4:} \quad \pi_4 = \frac{1}{16}, \quad q_4 = \frac{1}{2}, \quad \text{hence} \quad \pi_4 q_4 = \frac{1}{32}
\end{aligned}$$

Cell 1 has the highest product $\pi_1 q_1 = \frac{3}{16}$, making it the most strategic cell to search next.

The probability of the submarine being in a cell (π_k) and the probability of finding it provided it is there (q_k) are relevant when selecting the optimal cell to search. Intuitively, we want to maximize the information gained for each search operation. We can search either a cell where we believe there is a high chance the submarine may be and/or a cell where we can be almost sure we can cross out if the search team targeted it. Targeting the highest possible combination between location likelihood and search effectiveness is an expected result as it would deliver the most significant uncertainty reduction for the problem.

Although Cell 1 does not have the highest SEP (q_k), its higher inherent probability (π_1) makes it the most likely location to find the submarine when considering the combined metric $\pi_k q_k$. This approach balances the search effectiveness with the probability of locating the submarine in a given location, maximizing uncertainty reduction.