Problem 1:

Find the maximum entropy density \( f \), defined for \( x \geq 0 \), satisfying \( E[X] = \alpha_1 \), \( E[\ln X] = \alpha_2 \). That is, maximize \(-\int f \ln f \) subject to \( \int xf(x)dx = \alpha_1 \), \( \int (\ln x)f(x)dx = \alpha_2 \), where the integral is over \( 0 \leq x < \infty \). What family of densities is this?

Solution

The maximum entropy distribution subject to constraints

\[ \int xf(x)dx = \alpha_1 \] (1)

and

\[ \int (\ln x)f(x)dx = \alpha_2 \] (2)

is of the form

\[ f(x) = e^{\lambda_0 + \lambda_1 x + \lambda_2 \ln x} = cx^{\lambda_2}e^{\lambda_1 x} \] (3)

which is of the form of a Gamma distribution. The constants should be chosen so as to satisfy the constraints. We need to solve the following equations

\[ \int_0^\infty f(x)dx = \int_0^\infty cx^{\lambda_2}e^{\lambda_1 x} \, dx = 1 \] (4)

\[ \int_0^\infty xf(x)dx = \int_0^\infty cx^{\lambda_2+1}e^{\lambda_1 x} \, dx = \alpha_1 \] (5)

\[ \int_0^\infty (\ln x)f(x)dx = \int_0^\infty cx^{\lambda_2}e^{\lambda_1 x} \ln x \, dx = \alpha_2 \] (6)
Thus, the Gamma distributions $f(x) = \frac{1}{\Gamma(k)\theta} x^{k-1} e^{-x/\theta}$ with

$$E[X] = k\theta = \alpha_1 \quad E[\ln X] = \psi(k) + \ln(\theta) = \alpha_2$$

is the exponential family we want.

**Problem 2:**

What is the maximum entropy distribution $p(x, y)$ that has the following marginals?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$p_{11}$</td>
<td>$p_{12}$</td>
<td>$p_{13}$</td>
</tr>
<tr>
<td>2</td>
<td>$p_{21}$</td>
<td>$p_{22}$</td>
<td>$p_{23}$</td>
</tr>
<tr>
<td>3</td>
<td>$p_{31}$</td>
<td>$p_{32}$</td>
<td>$p_{33}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
</tr>
</tbody>
</table>

**Solution**

Given the marginal distributions of $X$ and $Y$, $H(X)$ and $H(Y)$ are fixed. Since the mutual information between $X$ and $Y$ are non-negative, we have

$$H(X, Y) = H(X) + H(Y) - I(X, Y) \leq H(X) + H(Y)$$

where the equality holds if and only if $X$ and $Y$ are independent. Hence, to maximize the entropy, $X$ and $Y$ should be independent, which requires $p_{X,Y}(x, y) = p_X(x) p_Y(y)$. Therefore, we have

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{7}$</td>
<td>$\frac{1}{7}$</td>
<td>$\frac{1}{7}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
</tr>
</tbody>
</table>

**Problem 3:**

Let $Y = X_1 + X_2$. Find the maximum entropy of $Y$ under the constraint $E[X_1^2] = P_1$, $E[X_2^2] = P_2$:

(a) If $X_1$ and $X_2$ are independent.

(b) If $X_1$ and $X_2$ are allowed to be dependent.

**Solution**

(a) If $X_1$ and $X_2$ are independent,

$$\text{Var}[Y] = \text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] \leq E[X_1^2] + E[X_2^2] = P_1 + P_2$$

where equality holds when $E[X_1] = E[X_2] = 0$. Thus we have

$$\max_{f(y)} h(Y) \leq \frac{1}{2} \log(2\pi e(P_1 + P_2))$$
where equality holds when $Y$ is Gaussian with zero mean, which requires $X_1$ and $X_2$ to be independent and Gaussian with zero mean.

(b) For dependent $X_1$ and $X_2$, we have

$$\text{Var}(Y) \leq \mathbb{E}[Y^2] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + 2\mathbb{E}[X_1X_2] \leq (\sqrt{P_1} + \sqrt{P_2})^2$$

where the first equality holds when $\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 0$, and the second equality holds when $Y$ is Gaussian with zero mean, which requires $X_1$ and $X_2$ to be Gaussian with zero mean and $X_2 = \sqrt{\frac{P_2}{P_1}}X_1$.

**Problem 4:**

We learned in the course that as long as the set of feasible means is open then every such mean can be realized by an element of the exponential family. In the following verify this explicitly (by not referring to the above statement for the following scenario).

(i) Let $\phi(x) = (x^2)$.

(ii) Let $\phi(x)$ consist of all elements $x_ix_j$, where $i$ and $j$ go from 1 to $K$.

**Solution**

Note that any covariance matrix can be realized by a Gaussian. In the same manner any feasible matrix of second moments can be realized by a Gaussian. And since Gaussians are elements of the exponential family the claim is proved.

**Problem 5:**

What is the maximum entropy distribution, call it $p(x,i)$, on $[0, \infty) \times \mathbb{N}$, both of whose marginals have mean $\mu > 0$. (I.e., in one axis the distribution is over the positive reals, whereas in the other one it is over the natural numbers.)

**Solution**

We know that entropy of a joint distribution is at most as large as the sum of the entropies of the marginals and we have equality if the two random variables are independent. Hence the solution will be the product distribution of two independent random variables. It hence suffices to find these two marginal distributions.

1. **Maximum entropy distribution on $[0, \infty)$ with mean $\mu$:** The answer is the exponential distribution with density $p(x) = \frac{1}{\mu}e^{-x/\mu}$. To see this note that the general form is an exponential distribution (i.e., a member of the exponential family) with the form $p(x) = e^{\theta x - A(\theta)}$ since the condition is that $\mathbb{E}[X] = \mu$, i.e., $\phi(x) = x$. The normalization constraint $\int_0^\infty p(x)dx = 1$ requires $e^{-A(\theta)} = -\theta$ and $\theta < 0$ so that $p(x) = -\theta e^{\theta x}$ for $\theta < 0$. And the mean constraint gives us that $\theta = -\frac{1}{\mu}$.

2. **In a similar manner the maximum entropy distribution on $\mathbb{N}$ with mean $\mu$ is equal to $p(i) = (1 - \frac{1}{\mu})^{i-1} \frac{1}{\mu} = (\mu - 1)^{i-1}/\mu^i$. Note that the maximum entropy of $p(i)$ is Geometric distribution, not Poisson distribution, since the support set is natural numbers $\{1, 2, 3, \ldots\}$.**

Therefore, the answer is $p(x,i) = e^{-x/\mu} \frac{(\mu-1)^{i-1}}{\mu^i}$.
Problem 6:

Let $P$ denote the zero-mean and unit-variance Gaussian distribution. Assume that you are given $N$ iid samples distributed according to $P$ and let $\hat{P}_N$ be the empirical distribution.

Let $\Pi$ denote the set of distributions with second moment $E[X^2] = 2$. We are interested in

$$\lim_{N \to \infty} \frac{1}{N} \log \Pr\{\hat{P}_N \in \Pi\} = -\inf_{Q \in \Pi} D(Q\|P).$$

(i) Determine $-\arg\inf_{Q \in \Pi} D(Q\|P)$, i.e., determine the element $Q$ for which the infinum is taken on.

(ii) Determine $-\inf_{Q \in \Pi} D(Q\|P)$.

Solution

We are looking for the $I$-projection of $P$ onto $\Pi$, call the result $Q$. Since $\Pi$ is a linear family with a single constraint on the expected value of $x^2$ we know that the density of the minimizing distribution has the form

$$q(x) = p(x)e^{\theta x^2 - A(\theta)}.$$  

If we insert $p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ this gives us

$$q(x) = e^{-x^2/2 + \theta x^2 - \tilde{A}(\theta)}.$$  

We recognize the right-hand side to be the density of a zero-mean Gaussian distribution and by assumption this distribution has second moment 2. Hence, the solution is a zero-mean Gaussian distribution with variance 2, i.e., $q(x) = \frac{1}{\sqrt{4\pi}} e^{-x^2/4}$. The asymptotic exponent is given by the KL distance between these two distributions. We have

$$D(q\|p) = \int \frac{1}{\sqrt{4\pi}} e^{-x^2/4} \log \frac{1}{\sqrt{4\pi}} e^{-x^2/4} dx$$

$$= \frac{1}{2} \log 2 + \int \frac{1}{\sqrt{4\pi}} e^{-x^2/2} \left[ -\frac{x^2}{4} + \frac{x^2}{2} \right] dx$$

$$= \frac{1}{2} \left( \log \frac{1}{2} + 1 \right) = \frac{1}{2} (-\log 2 + 1) \sim 0.153426.$$  

To summarize

1. $-\arg\inf_{Q \in \Pi} D(Q\|P)$ is given by $q(x) = \frac{1}{\sqrt{4\pi}} e^{-x^2/4}$.

2. $-\inf_{Q \in \Pi} D(Q\|P) = -0.153426$.  
