Problem Set 3 — *Due Friday, November 2, before class starts*

For the Exercise Sessions on Oct 26

Last name	First name	SCIPER Nr	Points

Rules : You are allowed and encouraged to discuss these problems with your colleagues. However, each of you has to write down her solution in her own words. If you collaborated on a homework, write down the name of your collaborators and your sources. No points will be deducted for collaborations. But if we find similarities in solutions beyond the listed collaborations we will consider it as cheating. Please note that EPFL has a VERY strict policy on cheating and you might be in BIG trouble. It is simply not worth it.

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 $\int x f(x) dx = \alpha_1$, $\int (\ln x) f(x) dx = \alpha_2$, where the integral is over $0 \leq x < \infty$. What family of densities is this?

Solution

The maximum entropy distribution subject to constraints

$$\int x f(x) dx = \alpha_1 \tag{1}$$

and

$$\int (\ln x) f(x) dx = \alpha_2 \tag{2}$$

is of the form

$$f(x) = e^{\lambda_0 + \lambda_1 x + \lambda_2 \ln x} = c x^{\lambda_2} e^{\lambda_1 x} \tag{3}$$

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

$$\int_0^\infty f(x) dx = \int_0^\infty c x^{\lambda_2} e^{\lambda_1 x} dx = 1 \tag{4}$$

$$\int_0^\infty x f(x) dx = \int_0^\infty c x^{\lambda_2 + 1} e^{\lambda_1 x} dx = \alpha_1 \tag{5}$$

$$\int_0^\infty (\ln x) f(x) dx = \int_0^\infty c x^{\lambda_2} e^{\lambda_1 x} \ln x dx = \alpha_2 \tag{6}$$

Thus, the Gamma distributions $f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$ with

$$E[X] = k\theta = \alpha_1 \qquad E[\ln X] = \psi(k) + \ln(\theta) = \alpha_2 \tag{7}$$

is the exponential family we want.

Problem 2:

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

x \ y	1	2	3	
1	p_{11}	p_{12}	p_{12}	$\frac{1}{2}$
2	p_{21}	p_{22}	p_{23}	$\frac{1}{4}$
3	p_{31}	p_{32}	p_{33}	$\frac{1}{4}$
	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	

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

$$H(X, Y) = H(X) + H(Y) - I(X, Y) \leq H(X) + H(Y) \tag{8}$$

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

x \ y	1	2	3	
1	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
2	$\frac{1}{6}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{4}$
3	$\frac{1}{6}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{4}$
	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	

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,

$$\text{Var}[Y] = \text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2] \leq E[X_1^2] + E[X_2^2] = P_1 + P_2 \tag{9}$$

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

$$\max_{f(y)} h(Y) \quad \leq \quad \frac{1}{2} \log(2\pi e(P_1 + P_2)) \tag{10}$$

where equality holds when Y is Gaussian with zero mean, which requires X_1 and X_2 to be independent and Gaussian with zeros mean.

(b) For dependent X_1 and X_2, we have

$$\mathrm{Var}(Y) \leq E[Y^2] = E[(X_1 + X_2)^2] = E[X_1^2] + E[X_2^2] + 2E[X_1 X_2] \leq (\sqrt{P_1} + \sqrt{P_2})^2 \tag{11}$$

where the first equality holds when $E[Y] = E[X_1] + E[X_2] = 0$, and the send equality holds when $X_2 = \sqrt{\frac{P_2}{P_1}} X_1$. Hence, $\max_{f(y)} h(Y) \leq \frac{1}{2}\log(2\pi e(\sqrt{P_1} + \sqrt{P_2})^2)$, where 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 $x_i x_j$, 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 μ: 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 μ 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 $\{1, 2, 3, \dots\}$.

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