Multivariate Methods

Homework #1 (2020-2021), Answers

Here we use Lagrange Multipliers to find maximum-entropy distributions. The common set-up for these problems is the following:

P *is a discrete probability distribution on a set of* N *values* $\{x_j\}$*: That is,* P_i *is the probability that a random draw chooses the value* x_i*. So* $P_i \geq 0$ *and* $\sum_{i=1}^{N} P_i = 1$.

The entropy of a probability distribution $H(P)$ *is defined as* $H(P) = -\sum_{i=1}^{N} P_i \log P_i$.

Q1: Find the distribution P *that maximizes* $H(P)$ *(subject to the constraint* $\sum_{i=1}^{N} P_i = 1$*).*

Our constraint is $\sum_{i=1}^{N} P_i = 1$. So we extremize $F(P,\lambda) = -\sum_{i=1}^{N} P_i \log P_i - \lambda \sum_{i=1}^{N} P_i$.

$$\frac{\partial}{\partial P_j} F(P,\lambda) = \frac{\partial}{\partial P_j}\left(-\sum_{i=1}^{N} P_i \log P_i - \lambda \sum_{i=1}^{N} P_i\right) = \frac{\partial}{\partial P_j}\left(-P_j \log P_j - \lambda P_j\right).$$
$$= -\log P_j - 1 - \lambda$$

So $\frac{\partial}{\partial P_j} F(P,\lambda) = 0$ is equivalent to $-\log P_j - 1 - \lambda = 0$. So all P_j are identical. The constraint $\sum_{i=1}^{N} P_i = 1$ then requires that $P_j = 1/N$.

Q2: Find the form of the distribution P *that maximizes* $H(P)$ *subject to a constraint on variance,* $\sum_{i=1}^{N} P_i x_i^{\,2} = V$ *-- you won't be able to solve for the values of the both Lagrange multipliers, but you can get close.*

We now have two constraints, $\sum_{i=1}^{N} P_i = 1$ and $\sum_{i=1}^{N} P_i x_i^{\,2} = V$. So we extremize

$$F(P,\lambda,\lambda_V) = -\sum_{i=1}^{N} P_i \log P_i - \lambda \sum_{i=1}^{N} P_i - \lambda_V \sum_{i=1}^{N} P_i x_i^2 .$$

$$\frac{\partial}{\partial P_j} F(P,\lambda,\lambda_V) = \frac{\partial}{\partial P_j}\left(-\sum_{i=1}^{N} P_i \log P_i - \lambda \sum_{i=1}^{N} P_i - \lambda_V \sum_{i=1}^{N} P_i x_i^2\right) = \frac{\partial}{\partial P_j}\left(-P_j \log P_j - \lambda P_j - \lambda_V P_j x_j^2\right).$$
$$= -\log P_j - 1 - \lambda - \lambda_V x_j^2$$

So $\frac{\partial}{\partial P_j} F(P,\lambda,\lambda_V) = 0$ means that $\log P_j = -1 - \lambda - \lambda_V x_j^2$, i.e., that $P_j = \exp\left(-1 - \lambda - \lambda_V x_j^2\right)$, which we can write more compactly as $P_j = K e^{-\lambda_V x_j^2}$. So the maximum-entropy distribution is a Gaussian, with the variance determined by V .