Derivation of Geometric Distribution from Maximum Entropy principle

Yaroslav Bulatov

February 7, 2005

Abstract

If we consider a set of discrete distributions over positive integers for which the first moment is known, the maximum entropy distribution of that set will be the geometric distribution.

Consider space of probability distributions $\mathcal{P} : \{0, 1, \ldots\} \to [0; 1]$. Now only consider a subset of that space for which $E[X] = m$ and maximize entropy. The is equivalent to solving the following problem:

$$arg \max_{\mathbf{p}\in\mathcal{P}} - \sum_{i=0}^{\infty} p_i \log p_i \tag{1}$$

subject to

$$\sum_i p_i = 1 \tag{2}$$

$$\sum_i i p_i = m \tag{3}$$

$$p_i \geq 0 \ \forall i \tag{4}$$

Actually the last constraint is rendundant because the solution will not lie on the boundary, so we can discard it.

We have two constraints, so introduce two Lagrange multipliers λ_0, λ_1 and form a Lagrangian

$$L(\mathbf{p}, \lambda_0, \lambda_1) = \sum_i p_i + \lambda_0 (\sum_i p_i - 1) + \lambda_1 (\sum_i i p_i - m) \tag{5}$$

To solve, find the gradient, and set to the $\mathbf{0}$ vector. This gives us an infinite system of equations, just enough to find unique solution.

$$\frac{\partial L}{\partial p_i}(\mathbf{p}, \lambda_0, \lambda_1) = -\log p_i - 1 + \lambda_0 + i\lambda_1 = 0 \tag{6}$$

rearranging terms, we get

$$p_i = \exp[\lambda_1 i] \exp[\lambda_0 - 1] \tag{7}$$

Substitute this into constraint (2) and get

$$\sum_i \exp[\lambda_1 i] = \exp[1 - \lambda_0] \tag{8}$$

Now we can substitute this into equation (7) to get rid of dependence on λ_0

$$p_i = \frac{\exp[\lambda_1 i]}{\sum_i \exp[\lambda_1 i]} \tag{9}$$

Substitute this into constraint (3) and get

$$\sum_i i \frac{\exp[\lambda_1 i]}{\sum_i \exp[\lambda_1 i]} = m \tag{10}$$

Rearranging terms we get

$$\sum_i i \exp[\lambda_1 i] = m \sum_i \exp[\lambda_1 i] \tag{11}$$

Now, remember the two geometric series formulae

$$\sum_{i=0}^{\infty} p^i = \frac{1}{1-p} \tag{12}$$

and

$$\sum_{i=0}^{\infty} i p^i = \frac{p}{(1-p)^2} \tag{13}$$

Apply the formulae to get

$$\frac{\exp[\lambda_1]}{(1 - \exp[\lambda_1])^2} = \frac{m}{1 - \exp[\lambda_1]} \tag{14}$$

Rearranging we get

$$\exp[\lambda_1] = \frac{m}{m+1} \tag{15}$$

Substituting back into equation (9) and aplying summation formula (12) once more we get

$$p_i = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^i \tag{16}$$

You can see the basic strategy is to solve for each variable in terms of others, eliminating them one by one.