# Common Probability Distributions

Qi Zhao

July 14, 2018

## 1. Introduction

For a view of probability from a machine learning perspective, there is one remaining important concept related to probability, which is conditional independence. The rules of probability are remarkably compact and simple. The concepts of marginalization [3], joint [2] and conditional probability [1], independence, and Bayesąŕ rule [5] will underpin many machine vision algorithms.

## 2. Description

To use these abstract rules for manipulating probabilities it will need to define some probability distributions. The choice of distribution Pr(x) that it use will depend on the domain of the data x that it are modeling (table 1).

Probability distributions such as the categorical and normal distributions are obviously useful for modeling visual data. When fitting probability models to data, we need to know how uncertain we are about the fit. This uncertainty is represented as a probability distribution over the parameters of the fitted model. So for each distribution used for modeling, there is a second distribution over the associated parameters (table 2). For example, the Dirichlet [4] is used to model the parameters of the categorical distribution. In this context, the parameters of the Dirichlet would be known as hyperparameters. More generally, the hyperparameters determine the shape of the distribution over the parameters of the original distribution.

| Data Type | Domain | Distribution |
|---|---|---|
| univariate, discrete, binary | $x \in 0, 1$ | Bernoulli |
| univariate, discrete, multi-valued | $x \in 1, 2, ..., K$ | categorical |
| univariate, continuous, unbounded | $x \in R$ | univariate normal |
| univariate, continuous, bounded | $x \in [0, 1]$ | beta |
| multivariate, continuous, unbounded | $x \in R_K$ | multivariate normal |
| multivariate, continuous, bounded, sums to one | $x = [x_1, x_2, ..., x_K]^T$ $x \in [0, 1]$ $\sum_{k=1}^{K} x_k = 1$ | Dirichlet |
| bivariate, continuous, $x_1$ unbounded, $x_2$ bounded below | x = [$x_1$ , $x_2$] $x_1 \in R$ $x_2 \in R^+$ | normal-scaled inverse gamma |
| multivariate vector **x** and matrix **X**, **x** unbounded, **X** square, positive definite | $x \in R^K$ $X \in R^{KXK}$ $z^T X z > 0 \,\forall\, z \in R$ | Knormal inverse Wishart |

Table 1. Common probability distributions: the choice of distribution depends on the type/domain of data to be modeled.

| Distribution | Domain | Parameters modeled by |
|---|---|---|
| Bernoulli | $x \in 0, 1$ | beta |
| categorical | $x \in 1, 2, ..., K$ | Dirichlet |
| univariate normal | $x \in R$ | normal inverse gamma |
| multivariate normal | $x \in R_K$ | normal inverse Wishart |

Table 2. Common distributions used for modeling (left) and their associated domains (center). For each of these distributions there is a second associated distribution over the parameters (right).

## References


[1] R. N. Aslin, J. R. Saffran, and E. L. Newport. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4):321–324, 1998. 1

[2] A. Fine. Hidden variables, joint probability, and the bell inequalities. *Physical Review Letters*, 48(5):291–295, 1982. 1

[3] R. S. Strauss and H. A. Pollack. Social marginalization of overweight children. *Archives of Pediatrics & Adolescent Medicine*, 157(8):746–752, 2003. 1

[4] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Sharing clusters among related groups: hierarchical dirichlet processes. In *Advances in Neural Information Processing Systems*, pages 1385–1392, 2005. 1

[5] R. A. Waller and D. B. Duncan. A Bayes rule for the symmetric multiple comparisons problem. *Journal of the American Statistical Association*, 64(328):1484–1503, 1969. 1