Probability and Statistics Review

Jonathan Navarrete

August 19, 2017

Distribution and Density functions

The cumulative distribution function (CDF) of a random variable X is

$$F_X(x) = P(X \leq x), \text{ where } x \in \mathbb{R}$$

In most situations, we will omit the X subscript and simply write $F(x)$ unless the support of F is unclear. A random variable X is continuous if F_X is a continuous function. A random variable X is discrete if F_X is a step function.

Discrete distributions can be specified by a probability mass function (pmf) $p_X(x) = P(X = x)$. If X id discrete, the CDF of X is

$$F_X(x) = P(X \leq x_k) = \sum_{i=1}^{k} p(p(x_1) + ... + p(x_k)$$

The continuous distributions do not have positive probability mass at any single point (theoretically). For continuous random variables X the probability density function (pdf) or density of X is $f_X(x) = \frac{d}{dx} F_X(x)$. Thus, the relationship between the pdf and CDF is

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

Some pdf/pmf's form a family of distributions. One of the most common distributions is the Normal distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2 \big/ 2} \text{ , } -\infty < x < \infty$$

The Normal distribution is part of the exponential family of distributions.

For more information, please see Related Distributions

Random Variables

Random variables are generated by a pdf/pmf. We say that a random variable X is distributed by $f_X(x)$ by

$$X \sim f_X(x)$$

A discrete random variable is one which may take on only a countable number of distinct values such as 0,1,2,. . . ,etc.. Discrete random variables are usually (but not necessarily) counts.

A continuous random variable is one which takes an infinite number of possible values. Continuous random variables are non-discrete measurements such as time, temperature, weight.

Expectation and Variance

The mean of a random variable X is the *expected value* or mathematical expectation of the random variable. If X is continuous with density $f(x)$, then

$$E[X] = \int_{-\infty}^{\infty} x \times f(x) dx$$

If X is discrete with pmf $f(x)$, then

$$E[X] = \sum_{x:p(x)>0} x \times f(x)$$

The variance of a random variable is

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

Likelihood

If there is a series of independent identically distributed (*iid*) random variables $\mathbf{x} = (x_1, ..., x_n)$, where $x_i \sim f(x|\theta)$, then the likelihood is given by

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i) = f(x_1) \times f(x_2) \times ... \times f(x_n)$$

The log-likelihood is then

$$l(\theta) = log(L(\theta|\mathbf{x})) = \sum_{i=1}^{n} log(f(x_i|\theta))$$

Maximum Likelihood Estimation

Maximizing likelihood is a method of estimating unknown population parameters. Given a random sample $\mathbf{x} = (x_1, ..., x_n)$, we calculate the log-likelihood $l(\theta)$ and solve for θ.

Other methods of parameter estimation include Method of Moments (MOM) and Bayesian Maximum A Posteriori (MAP).