Monte Carlo: Basics

Matthias Moulin

October 8, 2016

1 Characteristic values of a probability distribution

This section describes some important values including the mean, variance and standard deviation, to characterize or summarize a probability distribution.

Let $\mathrm{d}\mu(x)$ be the measure function:

$$\mathrm{d}\mu(x) := \begin{cases} p(x)\,\mathrm{d}x & \text{for continuous random variables} \\ \sum_{m=1}^{M} p_m \delta(x - x_m)\,\mathrm{d}x & \text{for discrete random variables} \end{cases}, \tag{1.1}$$

where p and p_m are a continuous and discrete probability distribution function, respectively.

The *mean* (a.k.a. *expected value* a.k.a. *expectation*) $\mathrm{E}[X]$ of the probability distribution with measure function $\mathrm{d}\mu(x)$ measures the expected value of a sample X. For a general function f, the *mean* $\mathrm{E}[f(X)]$ of the probability distribution with measure function $\mathrm{d}\mu(x)$ measures the expected value of a sample $f(X)$. Both are defined as:

$$\mathrm{E}[X] := \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} X_i = \int_{-\infty}^{+\infty} x\,\mathrm{d}\mu(x) \tag{1.2}$$

$$\mathrm{E}[f(X)] := \lim_{N \to +\infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}\mu(x), \tag{1.3}$$

where the X_i are sampled from the probability distribution with measure function $\mathrm{d}\mu(x)$ and are independent and identically distributed. Note that the mean is expressed in the same dimensions as the individual samples.

The *variance* $\mathrm{VAR}[X]$ of the probability distribution with measure function $\mathrm{d}\mu(x)$ measures the squared deviation of an individual sample X from its mean $\mathrm{E}[X]$. For a general function f, the *variance* $\mathrm{VAR}[f(X)]$ of the probability distribution with measure function $\mathrm{d}\mu(x)$ measures the squared deviation of individual sample $f(X)$ from its mean $\mathrm{E}[f(X)]$. Both are defined as:

$$\mathrm{VAR}[X] := \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \int_{-\infty}^{+\infty} (x - \mathrm{E}[X])^2 \mathrm{d}\mu(x) \tag{1.4}$$

$$\mathrm{VAR}[f(X)] := \mathrm{E}\left[(f(X) - \mathrm{E}[f(X)])^2\right] = \int_{-\infty}^{+\infty} (f(x) - \mathrm{E}[f(X)])^2 \mathrm{d}\mu(x), \tag{1.5}$$

where the X_i are sampled from the probability distribution with measure $\mathrm{d}\mu(x)$ and are independent and identically distributed. Note that the variance is expressed in the same but squared dimensions as the mean and individual samples.

The *standard deviation* $\mathrm{STD}[X]$ of the probability distribution with measure function $\mathrm{d}\mu(x)$ measures how an individual sample X deviates from its mean $\mathrm{E}[X]$. For a general function f, the *variance*

STD $[f(X)]$ of the probability distribution with measure function $d\mu(x)$ measures how an individual sample $f(X)$ deviates from its mean $E[f(X)]$. Both are defined as:

$$\text{STD}[X] := \sqrt{\text{VAR}[X]} \tag{1.6}$$

$$\text{STD}[f(X)] := \sqrt{\text{VAR}[f(X)]}, \tag{1.7}$$

where the X_i are sampled from the probability distribution with measure $d\mu(x)$ and are independent and identically distributed. Note that the standard deviation is expressed in the same dimensions as the mean and the individual samples.

The remainder of this text does not focus any more on the use of general functions f. However, it is trivial to include these.

The following properties for the mean and variance for independent random variables X, X_1, X_2, and scalar $\alpha \in \mathbb{R}^n$ can be useful:

$$E[X_1 + X_2] = E[X_1] + E[X_2] \tag{1.8}$$

$$E[X + \alpha] = E[X] + \alpha \tag{1.9}$$

$$E[\alpha \cdot X] = \alpha \cdot E[X] \tag{1.10}$$

$$\text{VAR}[X_1 + X_2] = \text{VAR}[X_1] + \text{VAR}[X_2] \tag{1.11}$$

$$\text{VAR}[X + \alpha] = \text{VAR}[X] \tag{1.12}$$

$$\text{VAR}[\alpha \cdot X] = \alpha^2 \cdot \text{VAR}[X]. \tag{1.13}$$

2 Estimators

In order to calculate the characteristic values of a probability distribution (defined in Section 1), we need to solve an integral (for which in general no tractable closed form expression exists) or alternatively use an infinite number of samples. In practice, we estimate these characteristic values with a finite number of samples N. We will denote *estimators* with a hat (e.g. $\widehat{A}$) or as $\langle B \rangle$ to stress the estimated quantity B.

An estimator $\langle E[X] \rangle$ for the expected value $E[X]$ of a sample X is defined as:

$$\langle E[X] \rangle := \frac{1}{N} \sum_{i=1}^{N} X_i. \tag{2.1}$$

An estimator $\langle \text{VAR}[X] \rangle$ for the variance $E[X]$ of a sample X is defined as:

$$\langle \text{VAR}[X] \rangle := \frac{1}{N} \sum_{i=1}^{N} (X_i - E[X])^2, \tag{2.2}$$

when the expected value of a sample X is known, or is defined as:

$$\langle \text{VAR}[X] \rangle := \frac{1}{N} \sum_{i=1}^{N} (X_i - \langle E[X] \rangle)^2, \tag{2.3}$$

when the expected value of a sample X is estimated as well with a different set of samples X_i, or is defined as:

$$\langle \text{VAR}[X] \rangle := \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \langle E[X] \rangle)^2, \tag{2.4}$$

where the expected value of of a sample X is estimated as well with the same set of samples X_i.

3 Error quantification

Since an estimator is obtained with only a finite number of samples (see Section 2), we need a way to quantify the 'error' between the estimator and the estimated quantity. Once, we are able to quantify the error in function of the number of samples, we can determine the number of samples needed for a given error.

The quantity $\widehat{X} - \mathrm{E}[X]$ where $\widehat{X}$ is an estimator for $\mathrm{E}[X]$, is called the *error*. The expected value of this error is called the *bias*, denoted as $\beta\left[\widehat{X}\right]$:

$$\beta\left[\widehat{X}\right] := \mathrm{E}\left[\widehat{X} - \mathrm{E}[X]\right] \tag{3.1}$$
$$= \mathrm{E}\left[\widehat{X}\right] - \mathrm{E}[X].$$

The *Mean Square(d) Error* (MSE) of an estimator $\widehat{X}$ is defined as:

$$\mathrm{MSE}\left[\widehat{X}\right] := \mathrm{E}\left[\left(\widehat{X} - X\right)^2\right] \tag{3.2}$$

$$= \mathrm{E}\left[\left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right] + \mathrm{E}\left[\widehat{X}\right] - X\right)^2\right]$$

$$= \mathrm{E}\left[\left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right]\right)^2 + 2 \cdot \left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right]\right) \cdot \left(\mathrm{E}\left[\widehat{X}\right] - X\right) + \left(\mathrm{E}\left[\widehat{X}\right] - X\right)^2\right]$$

$$= \mathrm{E}\left[\left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right]\right)^2\right] + 2 \cdot \mathrm{E}\left[\left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right]\right)\right] \cdot \mathrm{E}\left[\left(\mathrm{E}\left[\widehat{X}\right] - X\right)\right] + \mathrm{E}\left[\left(\mathrm{E}\left[\widehat{X}\right] - X\right)^2\right]$$

$$= \mathrm{E}\left[\left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right]\right)^2\right] + \underbrace{2 \cdot \left(\mathrm{E}\left[\widehat{X}\right] - \mathrm{E}\left[\widehat{X}\right]\right) \cdot \left(\mathrm{E}\left[\widehat{X}\right] - \mathrm{E}\left[\widehat{X}\right]\right)}_{=\,0} + \mathrm{E}\left[\widehat{X}\right]^2 - \mathrm{E}[X]^2$$

$$= \underbrace{\mathrm{E}\left[\left(\widehat{X} - \mathrm{E}\left[\widehat{X}\right]\right)^2\right]}_{\text{variance}} + \left(\underbrace{\mathrm{E}\left[\widehat{X}\right] - \mathrm{E}[X]}_{\text{bias}}\right)^2$$

$$= \mathrm{VAR}\left[\widehat{X}\right] + \beta\left[\widehat{X}\right]^2. \tag{3.3}$$

The *Root Mean Square(d) Error* (RMSE) of an estimator $\widehat{X}$ is defined as:

$$\mathrm{RMSE}\left[\widehat{X}\right] := \sqrt{\mathrm{MSE}\left[\widehat{X}\right]}. \tag{3.4}$$

To quantify the error in function of the number of samples N, we want to describe the behaviour of $\mathrm{RMSE}\left[\widehat{X}\right]$ in function of the number of samples N. We will estimate the quantity $\mathrm{RMSE}\left[\widehat{X}\right]$. In fact, we will use multiple independent estimators to characterize the distribution of the estimator. Once this distribution is known, we can validate candidate fitting curves on their feasibility.

For *unbiased* estimators $\widehat{X}$ the bias $\beta\left[\widehat{X}\right] \equiv 0$ or alternatively $\mathrm{E}\left[\widehat{X}\right] \equiv \mathrm{E}[X]$ which means that the estimator $\widehat{X}$ on average will be equal to the estimated quantity $\mathrm{E}[X]$ (independent of the number of samples N). Furthermore, $\mathrm{MSE}\left[\widehat{X}\right] \equiv \mathrm{VAR}\left[\widehat{X}\right]$ and $\mathrm{RMSE}\left[\widehat{X}\right] \equiv \sqrt{\mathrm{VAR}\left[\widehat{X}\right]}$.