

Notes about probability, and error estimation for sampled values

Cody L. Petrie

November 21, 2016

All of this material is being pulled from *Statistical Mechanics: Algorithms and Computations* by Werner Krauth.

1 Probability

1.1 Sum of random variables, convolution

- Random Variable: A variable that is subject to uncertainty.
- Random Variate: The actual value that a random variable takes on. For example, if you toss dice X may be the random variable with $1/6$ probability of being any number between 1 and 6. Not if you toss a dice and get 2, then 2 is the random variate, because it is the value that the random variable takes on.

If you have a random number generator that generates N **independent** numbers, $k_1 \dots k_N$ (random variates of the random variables $\{\xi_1 \dots \xi_N\}$) each with probability $\pi_i = \pi(k_i)$, then the probability of getting a particular string of number is

$$\pi(k_1 \dots k_N) = \pi(k_1) \dots \pi(k_N). \quad (1)$$

Let's say that these values can be either 1 (hit) or 0 (nonhit). This means that the sum random variable (the number of hits) is ξ , and can be $0, \dots, N$. Since there are $\binom{N}{k} = \frac{N!}{k!(N-k)!}$ ways to choose the k values with N trials, the probability that the number of hits, ξ , will be k is

$$\pi_k = \binom{N}{k} \theta^k (1 - \theta)^{N-k}, \quad (2)$$

where θ is the probability of a hit. Then you can see that the probability of k hits after $N + 1$ trials is π'_k

$$\pi'_k = \pi_k \cdot (1 - \theta) + \pi_{k-1} \cdot \theta. \quad (3)$$

This looks like a convolution and in continuous form can be written as

$$\pi'(x) = \int_{x-1}^x dy \pi(y) \pi^1(x - y) \quad (4)$$

1.2 Mean value and variance

- Second moment: This is the variance and is special because it is additive.
- Independent random variables: variables that satisfy $\pi(x_1, x_2) = \pi(x_1)\pi(x_2)$.
- Chebyshev's Inequality: $P(|x - \langle x \rangle| > \epsilon) < \frac{\text{Var}(\xi)}{\epsilon^2}$

The expectation value (mean) of discrete and continuous variables ξ can be written respectively as

$$\langle \xi \rangle = \sum_k k \pi_k, \quad (5)$$

$$\langle \xi \rangle = \int dx x \pi(x). \quad (6)$$

Also mean value of the sum of two random variables is the sum of the means (even when the variables are not independent).

$$\langle \xi_1 + \xi_2 \rangle = \int dx_1 dx_2 (x_1 + x_2) \pi(x_1, x_2) = \langle \xi_1 \rangle + \langle \xi_2 \rangle \quad (7)$$

The variance (The second, and probably most talked about moment) is given by the squared deviation from the mean,

$$\text{Var}(\xi) = \langle (\xi - \langle \xi \rangle)^2 \rangle = \langle \xi^2 - 2\xi \langle \xi \rangle + \langle \xi \rangle^2 \rangle = \langle \xi^2 \rangle - \langle \xi \rangle^2, \quad (8)$$

and the standard deviation is the square root of this value, $\sigma = \sqrt{\text{Var}(\xi)}$.

I am having trouble showing, but it turns out that for **independent** variables the variance of the sum of random variables is the sum of their variances.

$$\text{Var}(\xi_1 + \dots + \xi_N) = \text{Var}(\xi_1) + \dots + \text{Var}(\xi_N). \quad (9)$$

1.3 The central limit theorem

Any probability distribution can be shifted by the mean and scaled to have a variance of one. For example,

$$\pi_{\text{resc}}(x) = \sigma \pi(\sigma x + \langle \xi \rangle), \quad (10)$$

where the argument is $y = \sigma x + \langle \xi \rangle$, and the variable ξ takes on the value y with probability $\pi(y)$. Then the variable ξ_{resc} takes on the values $x = (y - \langle \xi \rangle)/\sigma$ with probability $\pi_{\text{resc}}(x)$. The **central limit theorem** then states that the rescaled random variable is Gaussian in the limit that $N \rightarrow \infty$, if ξ is the sum of independent random variables, $\xi = \xi_1 + \dots + \xi_N$ of finite variance.

2 Error Estimation

Whether the random variables are independent or dependent the mean value is going to be the same. However, for dependent random variates the error will be an under estimate because random variates will be closer to the previous variate, as in a Markov chain.

2.1 Independent random variables

Calculating the mean value is straight forward.

$$\mu = \frac{1}{N} \sum_{i=1}^N \xi_i \quad (11)$$

The error is then given by the standard deviation (square root of the variance, or second moment).

$$\sigma = \frac{1}{\sqrt{N-1}} \sqrt{\left[\sum_{j=1}^N \left(\xi_j - \frac{1}{N} \sum_{i=1}^N \xi_i \right)^2 \right]} \quad (12)$$

$$= \frac{1}{\sqrt{N-1}} \sqrt{\frac{1}{N} \sum_i \xi_i^2 - \left(\frac{1}{N} \sum_i \xi_i \right)^2} \quad (13)$$

$$\approx \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_i \xi_i^2 - \left(\frac{1}{N} \sum_i \xi_i \right)^2} \quad (14)$$