

Notes on Stats

April 6, 2024

Contents

1	Dynamic Properties of the Monte Carlo Method in Statistical Physics	2
2	Central Limit Theorem	2
3	Gaussians and the Error Function	2
4	Censored Data Models	7

1 Dynamic Properties of the Monte Carlo Method in Statistical Physics

These are notes from the paper "Dynamic Properties of the Monte Carlo Method in Statistical Physics" by H. Muller-Krumbhaar and K Binder, Journal of Statistical Physics, Vol 8, No. 1, 1973.

References 5, 6, 27, and 28 mention processes for estimating statistical errors.

However, the processes mentioned in ref 28 disregard the knowledge that one can draw from transition probabilities about the well-defined correlations between subsequent configurations.

2 Central Limit Theorem

A collection of random variables is independent and identically distributed (IID) if each random variable has the same probability distribution as the others and all are mutually independent.

Another way of thinking about an IID set of random variables is as a random sample.

If an arbitrarily large number of samples, each involving multiple observations (data points), were separately used in order to compute one value of a statistic (such as, for example, the sample mean or sample variance) for each sample, then the sampling distribution is the probability distribution of the values that the statistic takes on.

The central limit theorem establishes that, in many situations, for independent and identically distributed random variables, the sampling distribution of the standardized sample mean tends towards the standard normal distribution even if the original variables themselves are not normally distributed.

Normal distributions are often used to represent real-valued random variables whose distributions are not known because of the central limit theorem. One example of IID data could be measurement errors.

3 Gaussians and the Error Function

The error function was first formally introduced in On a class of definite integrals by J.W.L. Glaisher B.A. F.R.A.S. F.C.P.S.. (We found that link in the wikipedia page talking about the Error function Error function, it was reference number 3.)

Some great references we recommend watching and working through are

1. The impossible integral of $e^{\hat{x}^2}$ & the error function by blackpenredpen
2. Integral of $e^{\hat{x}^2}$ & the Imaginary Error Function by blackpenredpen

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

This is the definite integral version of a Gaussian integral. Since $\int_{-\infty}^{\infty} x^{-x^2} dx = \sqrt{\pi}$ and since the definite integral we use to define the error function only goes from 0 to x (only half the domain of the Gaussian), that's why we normalize the $\operatorname{erf}(x)$ with a $\sqrt{\pi}/2$.

To draw the horizontal asymptotes of $\operatorname{erf}(x)$, note that the domain is $(-\infty, \infty)$, so we need to take the limits of x as it tends to positive and negative infinity.

$$\begin{aligned} \lim_{x \rightarrow -\infty} \operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^{-\infty} e^{-t^2} dt \\ &= -\frac{2}{\sqrt{\pi}} \int_{-\infty}^0 e^{-t^2} dt \\ &= -1 \end{aligned}$$

Similarly, $\lim_{x \rightarrow \infty} \operatorname{erf}(x) = 1$.

The derivative is also quite easy to get since we are taking the derivative of an integral and thus the fundamental theorem of calculus is all we need,

$$\operatorname{erf}'(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}$$

From the derivative we can also tell that $\operatorname{erf}(x)$ has no critical numbers and as a matter of fact, the derivative is always non-negative.

And since we are talking about critical points, we can also go and take a look at the inflection points by computing the second derivative,

$$\operatorname{erf}''(x) = -2x \frac{2}{\sqrt{\pi}} e^{-x^2}$$

From this expression we can tell that $x = 0$ is an inflection point, and that for all $x < 0$ $\operatorname{erf}(x)$ is concave down, and for $x > 0$ $\operatorname{erf}(x)$ is concave up.

Now, what about $\int e^{x^2} dx$, here we could do a variable transformation,

$$\begin{aligned}
\int e^{x^2} dx &= \int e^{-(-x^2)} dx \\
&= \int e^{-(ix)^2} dx \rightarrow \begin{bmatrix} u = ix & x = \frac{1}{i}u = \frac{1}{i}i u = -iu \\ du = i dx & dx = -i du \end{bmatrix} \\
&= \int e^{-(u)^2} (-i) du \\
&= -i \int e^{-u^2} du \\
&= -i \frac{2}{\sqrt{\pi}} \operatorname{erf}(u) \\
&= -i \frac{2}{\sqrt{\pi}} \operatorname{erf}(ix) = \operatorname{erfi}(x)
\end{aligned}$$

There is one more property to mention before we continue looking into variable transformations and that is that $\operatorname{erf}(-z) = -\operatorname{erf}(z)$, the error function is an odd function. This directly results from the fact that the integrand e^{-t^2} is an even function (the antiderivative of an even function which is zero at the origin is an odd function and vice versa).

Glasser's Master Theorem

We originally learned of it thanks to A masterful theorem for integration by Michael Penn. But the theorem goes as follows:

$$\int_{-\infty}^{\infty} f\left(x - \frac{a}{x}\right) dx = \int_{-\infty}^{\infty} f(x) dx$$

for all $a > 0$.

If we try the most obvious u-substitution possible $u = x - \frac{a}{x}$, we can rewrite it as $ux = x^2 - a$ or $x^2 - ux - a = 0$. If we apply the quadratic equation $x = \frac{1}{2}(u \pm \sqrt{u^2 + 4a})$. This is also one of the reason why $a > 0$ - to prevent a negative number inside of the square root. Consequently, $dx = \frac{1}{2} \left(1 \pm \frac{u}{\sqrt{u^2 + 4a}}\right) du$.

If we take the "+" sign, of x , then all of our values will be positive. And if we take the "-" sign, then all of our values will be negative.

If we now do the u-substitution and take into account the nature of x depending on what sign we take, then,

$$\begin{aligned}
\int_{-\infty}^{\infty} f\left(x - \frac{a}{x}\right) dx &= \int_{-\infty}^0 f\left(x - \frac{a}{x}\right) dx + \int_0^{\infty} f\left(x - \frac{a}{x}\right) dx \\
&= \frac{1}{2} \int f(u) \left(1 - \frac{u}{\sqrt{u^2 + 4a}}\right) du + \frac{1}{2} \int f(u) \left(1 + \frac{u}{\sqrt{u^2 + 4a}}\right) du
\end{aligned}$$

Now let's figure out the bounds for integration. $\lim_{x \rightarrow -\infty} u = x - \frac{a}{x} = -\infty$ and $\lim_{x \rightarrow 0^-} u = x - \frac{a}{x} = \infty$ (because $x \rightarrow 0^-$ "x approaches 0 from the left.")

Similarly, in the second integral, x will approach 0 from above, $x \rightarrow 0^+$, so $\lim_{x \rightarrow 0^+} u = x - \frac{a}{x} = -\infty$. And $\lim_{x \rightarrow \infty} u = x - \frac{a}{x} = \infty$. So we have,

$$\begin{aligned} \int_{-\infty}^{\infty} f\left(x - \frac{a}{x}\right) dx &= \int_{-\infty}^0 f\left(x - \frac{a}{x}\right) dx + \int_0^{\infty} f\left(x - \frac{a}{x}\right) dx \\ &= \frac{1}{2} \left[\int_{-\infty}^{\infty} f(u) \left(1 - \frac{u}{\sqrt{u^2 + 4a}}\right) du + \int_{-\infty}^{\infty} f(u) \left(1 + \frac{u}{\sqrt{u^2 + 4a}}\right) du \right] \\ &= \int_{-\infty}^{\infty} f(u) du \\ &= \int_{-\infty}^{\infty} f(x) dx \end{aligned}$$

$$\begin{aligned} \frac{\pi}{b} &= \frac{1}{b} \left(\lim_{x \rightarrow \infty} \arctan\left(\frac{x}{b}\right) - \lim_{x \rightarrow -\infty} \arctan\left(\frac{x}{b}\right) \right) \\ &= \frac{1}{b} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) \end{aligned}$$

From calculus,

$$\frac{d}{dx} \arctan x = \frac{1}{x^2 + 1}$$

So if we had,

$$dy = \frac{dx}{x^2 + b^2} = \frac{dx}{b^2 \left(\left(\frac{x}{b}\right)^2 + 1 \right)}$$

If we choose $u = x/b$ so that $du = dx/b$, or $dx = bdu$, then

$$\begin{aligned} \int dy &= \int \frac{1}{b^2} \frac{bdu}{u^2 + 1} = \int \frac{1}{b} \frac{du}{u^2 + 1} \\ &= \frac{1}{b} \arctan u + C \\ &= \frac{1}{b} \arctan\left(\frac{x}{b}\right) + C \end{aligned}$$

Differentiating the inverse tangent function

Anyway, pluting everything together,

$$\begin{aligned} \frac{\pi}{b} &= \frac{1}{b} \left(\lim_{x \rightarrow \infty} \arctan\left(\frac{x}{b}\right) - \lim_{x \rightarrow -\infty} \arctan\left(\frac{x}{b}\right) \right) \\ &= \int_{-\infty}^{\infty} \frac{dx}{x^2 + b^2} \end{aligned}$$

If we use Glasser's theorem,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{dx}{x^2 + b^2} &= \int_{-\infty}^{\infty} \frac{dx}{\left(x - \frac{a}{x}\right)^2 + b^2} \\
&= \int_{-\infty}^{\infty} \frac{dx}{x^2 + \left(\frac{a}{x}\right)^2 - 2a + b^2} \\
&= \int_{-\infty}^{\infty} \frac{x^2 dx}{x^4 + (b^2 - 2a)x^2 + a^2} \\
&= \frac{\pi}{b}
\end{aligned}$$

Example of how the above can be used: consider

$$\int_{-\infty}^{\infty} \frac{2x^2 dx}{x^4 + 2x^2 + 5}$$

By comparison with our previous result we see that $a = \sqrt{5}$ and so $b^2 - 2a = 2$. This system of equation can be solved by writing the latter equation in terms of b , that is $b = \sqrt{2 + 2\sqrt{5}}$. Hence,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{2x^2 dx}{x^4 + 2x^2 + 5} &= \frac{2\pi}{\sqrt{2 + 2\sqrt{5}}} \\
&= \pi \sqrt{\frac{2}{1 + \sqrt{5}}} \\
&= \frac{\pi}{\sqrt{\phi}}
\end{aligned}$$

Where ϕ is the golden ratio.

4 Censored Data Models

Censored data models are missing data models where densities are not sample directly. And like any sort of model, we want to obtain estimates and make inferences.

In a typical statistical model, we would observe random variables Y_1, \dots, Y_n , drawn from a populaion with distribution $f(y|\theta)$. The distribution of the sample would then be given by $\prod_{i=1}^n f(y_i|\theta)$. Inference about θ would be based on this distribution.

In many situations we have to deal with Censored random variables; that is, rather than observing Y_i we may observe $\min\{Y_i, \bar{u}\}$, where \bar{u} is a constant.

Several types of censoring can be categorized by their relation with an underlying, unobserved, model $Y_i \sim f(y_i|\theta)$.

1. Given random variables Y_i , which are, for instance, times of observation, the actual observations are $Y_i^* = \min\{Y_i, \bar{u}\}$, where \bar{u} is the maximal observation duration or some truncation point.
2. The original variables Y_i are kept in the sample with probabilities $p(y_i)$ and the numver of censored variables is either known or unknown.
 - (a) This concept introduces a probabilistic approach to censoring, where each observation Y_i has a certain probability $p(y_i)$ of being included (i.e., not censored) in the sample. This probability can depend on the value of Y_i itself, introducing a form of randomness to which observations are censored.
 - (b) This contrasts with traditional views of censoring, where the censoring mechanism might be considered non-random or fixed.
 - (c) Known or unknown number of censored variables refers to whether the total count of censored observations in the sample is a known quantity or not. In some studies, the design might ensure that a specific number of observations will be censored (known), whereas in others, the number might vary randomly (unknown).
3. The variables Y_i are associated with auxilliary variables $X_i \sim g$ such that $y_i^* = h(y_i, x_i)$ is the observation. Typically $h(y_i, x_i) = \min(y_i, x_i)$. The fact that truncation ocurred, namely the variable $\mathbb{I}_{Y_i > X_i}$, may be either known or unknown.

$X \sim N(\theta, \sigma^2)$ and $Y \sim N(\mu, \tau^2)$

the variables $Z = X \wedge Y = \min(X, Y)$ is distributed as

$$\left[1 - \Phi\left(\frac{z - \theta}{\sigma}\right)\right] \times \frac{1}{\tau} \phi\left(\frac{z - \mu}{\tau}\right) + \left[1 - \Phi\left(\frac{z - \mu}{\tau}\right)\right] \times \frac{1}{\sigma} \phi\left(\frac{z - \theta}{\sigma}\right)$$

where ϕ is the density of the normal $N(0, 1)$ distribution and Φ is the corresponding CDF.

The density of $Z = \min(X, Y)$ is derived from considering the ways in which Z could take a particular value z .

First, $X < Y$ and $X = z$. Here the density at z is given by $\phi\left(\frac{z-\theta}{\sigma}\right)$ (ϕ is the PDF of X), and Y must be greater than $X = z$, hence $1 - \Phi\left(\frac{z-\mu}{\tau}\right)$ (Φ is the CDF of Y).

Second, $X > Y$ and $Y = z$. So for Z to equal z , Y must equal z , which is given by the PDF of Y at z , $\phi\left(\frac{z-\mu}{\tau}\right)$. And X must be greater than z , the probability of which is $1 - \Phi\left(\frac{z-\theta}{\sigma}\right)$.

The scaling factors of σ^{-1} and τ^{-1} come because when you standardize X and Y the standard normal density function ϕ is used, but you must also scale by the standard deviation because the change of variables affects the density's height. This ensures that the total area under the PDF remains 1, preserving the probability density function's properties.