

Crystal Ball Gazing: Mathematical models for chasing returns

Thomas P. Steele

May 27, 2025

Contents

1	Fundamentals	2
1.1	Stochastic Models	2
1.2	Bayes Theorem and Bayesian Inference	4
1.3	Bayesian Linear Regression	5
1.4	Financial Random Walks	5
1.5	Expectations of stochastic models	5

Chapter 1

Fundamentals

1.1 Stochastic Models

A model in general takes some input and gives some output,

$$f(x) = y$$

A stochastic model is the same, however, every single time we evaluate the model we get a different answer.

To get a feel for stochastic modelling, we will start with a basic stochastic model. Suppose we are studying some phenomena and collect a bunch of observations. We will call the set of observations Y and each observation y_i . Our goal is to make a stochastic model which mimics our set of observations,

$$y_i = \mu + \varepsilon_i$$

you can think of ε as the *error*, so our model is that each of our observations has a value μ plus some error. For this type of model we must make an assumption on how our error is distributed, the easiest and often best assumption is that our error is normally distributed with mean 0 and some standard deviation σ , $\varepsilon_i \sim \mathcal{N}(0, \sigma)$. Our model then becomes,

$$y_i \sim \mathcal{N}(\mu, \sigma),$$

meaning, our observations are randomly distributed by the normal distribution with mean μ and standard deviation σ . To actually learn anything from this model, we must first learn about the likelihood function.

The likelihood function is the total probability of our observations given our observations. What it tells us is the total probability of getting a set of obser-

vations given a model. The full form of the likelihood function is,

$$L(y_i | \theta) = \prod_{i=1}^n \pi(y_i | \{y_j\}_{j \neq i}^n, \theta).$$

Here θ is all of our model parameters, in this case μ and σ . A factor we must consider about our data is whether it is independent and identically distributed (iid) or not. iid data means that all of our observations are completely independent from one another, we will worry about non-iid data later. Here assuming our data is iid the likelihood becomes,

$$\begin{aligned} L(y_i | \theta) &= \prod_{i=1}^n \pi(y_i | \theta), \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right), \\ &= \frac{1}{\sigma^n(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right). \end{aligned}$$

The parameters which are the most likely to give us our observations maximises the likelihood function, so to infer the value of our parameters, we can find for what values is the likelihood function maximised. Before we get started on that, let's talk about the log-likelihood. The log-likelihood is exactly the same, but, wildly easier to differentiate,

$$\begin{aligned} \ell(Y | \theta) &= \ln(L(Y | \theta)) \\ &= -\frac{n}{2} \ln(2\pi) - n \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \end{aligned}$$

To find the best value for μ we solve,

$$\begin{aligned} 0 &= \frac{\partial \ell(Y | \theta)}{\partial \mu}, \\ 0 &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu), \\ \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n y_i, \end{aligned}$$

and for σ ,

$$\begin{aligned} 0 &= \frac{\partial \ell(Y | \theta)}{\partial \sigma}, \\ 0 &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (y_i - \mu)^2, \\ \hat{\sigma} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2}. \end{aligned}$$

We give our estimates $\hat{\mu}$ and $\hat{\sigma}$ to signify that these are estimated values. If these formulas look familiar, they should, these are the mean and standard deviation formulas, the mean in particular you probably use far more often than you realise.

1.2 Bayes Theorem and Bayesian Inference

Summation rule

$$\sum_{i=1}^n \Pr(x, y_i) = \Pr(x) \quad (1.1)$$

Product Rule

$$\Pr(A \cap B) = \Pr(A) \Pr(B | A) \quad (1.2)$$

Bayes theorem comes from thinking about the product rule being commutative,

$$\Pr(A \cap B) = \Pr(B \cap A) \quad (1.3)$$

$$\Pr(A) \Pr(B | A) = \Pr(B) \Pr(A | B) \quad (1.4)$$

$$\Pr(A | B) = \frac{\Pr(A) \Pr(B | A)}{\Pr(B)} \quad (1.5)$$

The usual form of Bayes' theorem we will see when doing Bayesian Inference is,

$$\pi(\theta | X) = \frac{L(X | \theta) \pi(\theta)}{\int_{\theta \in \Theta} L(X | \theta) \pi(\theta) d\theta} \quad (1.6)$$

The denominator comes from Bayes Theorem,

$$\begin{aligned} \pi(X) &= \int_{\theta \in \Theta} \pi(X \cap \theta) d\theta \\ \pi(X \cap \theta) &= L(X | \theta) \pi(\theta) \\ \pi(X) &= \int_{\theta \in \Theta} L(X | \theta) \pi(\theta) d\theta \end{aligned}$$

1.3 Bayesian Linear Regression

1.4 Financial Random Walks

$$S_{t+1} = S_t + \mu\Delta t S_t + \sigma\sqrt{\Delta t} S_t Y_i$$

$$S_{t+1} = S_t \left(1 + \mu\Delta t + \sigma\sqrt{\Delta t} Y_i\right)$$

$Y_i \sim \mathcal{N}(0, 1)$ so $\sqrt{\Delta t} Y_i = \mathcal{N}(0, t)$, and by definition the Wiener process $W_t = \mathcal{N}(0, t)$.

$$S_{t+\Delta t} = S_t (1 + \mu\Delta t + \sigma W_{\Delta t})$$

taking the limit $\Delta t \rightarrow dt$ we get the differential form of a random walk,

$$S(t + dt) = S(t) (1 + \mu dt + \sigma dW)$$

One way to solve this is to approximate the movement with a discrete time step. Combining multiple steps we get the equation,

$$S_{t+n} = S_t \prod_{i=1}^n (1 + \mu\Delta t + \sigma W_t)$$

taking the log we find,

$$\ln([S_{t+n}]) = \ln([S_t]) + \sum_{i=1}^n \ln([1 + \mu\Delta t + \sigma W_t])$$

if μ and σ are small then we can approximate this as

$$\ln([S_{t+n}]) = \ln([S_t]) + \sum_{i=1}^n \mu\Delta t + \sigma W_t$$

this is essentially a normal random walk where we have a deterministic drift μ and variance given by the Wiener process.

1.5 Expectations of stochastic models

If we presume there is no drift we assume there is no gains.