Methodology for Deriving Option-Implied Probability Distributions

Jan Oberwahrenbrock

July 2025

1 Defining a Suitable Representation for Options

Given an underlying asset (e.g., BTC) with current price c, we seek a practical representation for a European call option with strike s and time to expiration t. We will focus on European call options; however, the methodology can be applied analogously to European put options. We represent each option by the tuple (d, t), where

$$d = s - c.$$

As we will see later, this choice of representation offers some advantages. Consequently, a set of options can be described as a set of such tuples.

2 Present Value of an Option

The function $V\colon (d, t) \mapsto \mathbb{R}_{>0}$ returns the present (fair) value of an option specified by the tuple (d, t). It is defined as

$$V(d,t) = \underbrace{e^{-r\,t}}_{\substack{\text{discount factor}\\ \text{(time value of money)}}} \times \underbrace{\mathbb{E}_{p_t}\big[\max\{x-d,0\}\big]}_{\substack{\text{expected payoff}\\ \text{at expiry}}} = e^{-r\,t}\int_{-c}^{\infty} \max\{x-d,0\}\, p_t(x)\,\mathrm{d}x.$$

where

- r is the risk-free interest rate,
- x denotes the change in the underlying asset's price,
- $p_t(x)$ is the assumed probability density for the price change x over the remaining time t.

Since the underlying price cannot fall below zero, we restrict the lower integration limit to $x = -c$. For small t, the discount factor satisfies $e^{-r\,t} \approx 1$.

3 Deriving Implied Probability Distribution from Present Value

Assuming a small, fixed maturity t (so that $e^{-r\,t} \approx 1$), the present-value function simplifies to

$$V(d) = \int_{-c}^{\infty} \max\{x - d, 0\}\; p_t(x)\,\mathrm{d}x.$$

$$= \int_{d}^{\infty} (x - d) p_t(x)\,\mathrm{d}x$$

$$= \int_{d}^{\infty} x p_t(x)\,\mathrm{d}x - \int_{d}^{\infty} d p_t(x)\,\mathrm{d}x$$

$$= \int_{d}^{\infty} x p_t(x)\,\mathrm{d}x - d\int_{d}^{\infty} p_t(x)\,\mathrm{d}x$$

Differentiating the simplified present-value function $V(d)$ with respect to d yields

$$\frac{\mathrm{d}V}{\mathrm{d}d} = -\int_{d}^{\infty} p_t(x)\,\mathrm{d}x = -(1 - \int_{-c}^{d} p_t(x)\,\mathrm{d}x) = \int_{-c}^{d} p_t(x)\,\mathrm{d}x - 1$$

In the second equality, we have applied the complement rule for probabilities, Recall the Fundamental Theorem of Calculus:

$$F(x) = \int_{a}^{x} f(t)\,\mathrm{d}t \quad \Longrightarrow \quad \frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x).$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}d}\frac{\mathrm{d}V}{\mathrm{d}d} = \frac{\mathrm{d}^2V}{\mathrm{d}d^2} = p_t(d)$$

In other words, by taking the second derivative of the present-value function with respect to d, we directly recover the implied probability density $p_t(d)$ for the underlying price change over the remaining time t.

4 Market Prices and Present Value

In option pricing we distinguish three layers of probability distributions:

1. **Physical (real-world) distribution.** There exists an underlying "physical" distribution $p_{\text{phys}}(x)$ for the true price changes, but it is inherently unknowable.

2. **Market participants' model of the physical distribution.** To price options, traders and quants adopt a model $p_{\text{model}}(x)$ meant to approximate $p_{\text{phys}}(x)$.

3. **Seller premium and market-implied distribution.** On top of their model price, sellers demand a risk premium. As a result, the observed market price $P(d,t)$ exceeds the model-based present value $V(d,t)$:

$$P(d,t) \;\geq\; V(d,t).$$

By differentiating $P(d,t)$ twice with respect to d, we obtain the market-implied density

$$p_{\text{implied}}(d) = \frac{\partial^2 P(d,t)}{\partial d^2},$$

which tends to be broader than the original model $p_{\text{model}}(x)$. This greater dispersion is reflected in the empirical observation that, on average, implied volatility exceeds realized (historical) volatility.

Because the true physical distribution $p_{\text{phys}}(x)$ cannot be known and the sellers' risk premium cannot be easily extracted, the market-implied density $p_{\text{implied}}(x)$ remains our most practical proxy for the market participants' model distribution $p_{\text{model}}(x)$.

5 Constructing a Twice-Differentiable Price Function

Market quotes for a given expiration date are available at only a few dozen discrete strikes. For example, on Deribit—the largest BTC-options exchange—adjacent strikes near the current price c may be \$500 apart, while strikes farther out can differ by several thousand dollars. Such a discrete set of prices does not define a continuous, twice-differentiable function. Therefore, we require a method to interpolate $P(d,t)$ into a smooth C^2 function.

5.1 Generating Additional Data Points

We exploit the option representation (d,t) with $d = s - c$ to obtain more observations over time. After fetching the current option quotes, we pause for a few seconds and fetch again. Whenever the underlying price c has moved—as is common for a liquid asset like BTC—we receive a new set of strikes. Since the implied density $p_t(x)$ is unlikely to change over short intervals (minutes) when t spans days and no major news occur, repeating this process for several minutes yields hundreds of distinct data points.

5.2 Fitting a Smooth Proxy via Constrained Splines

To construct a C^2 approximation $\tilde{P}(d,t)$ of the discrete quotes $P(d_i,t)$, we fit a spline of degree at least three on a sequence of support points $\{p_n\}$. On each interval $[p_n, p_{n+1}]$, let $f_n(d)$ be the local polynomial. At every interior support point p_n, enforce continuity of value and its first two derivatives:

$$f_{n-1}(p_n) = f_n(p_n), \quad f'_{n-1}(p_n) = f'_n(p_n), \quad f''_{n-1}(p_n) = f''_n(p_n).$$

This guarantees $\tilde{P}(d,t) \in C^2$.

Since $\tilde{P}''(d)$ represents a probability density, we further impose shape constraints:

$$\tilde{P}''(d) > 0, \quad \tilde{P}'''(d) > 0 \text{ for } d < c, \quad \tilde{P}'''(d) < 0 \text{ for } d > c,$$

which imply $\tilde{P}''(c)$ is its global maximum.

The coefficients are obtained by minimizing the sum of squared errors,

$$\sum_i \big(\tilde{P}(d_i,t) - P(d_i,t)\big)^2,$$

subject to the continuity and shape constraints. The result is a smooth, arbitrage-free proxy $\tilde{P}(d,t)$ from which we recover the implied density via

$$p_t(d) \;=\; \frac{\partial^2 \tilde{P}(d,t)}{\partial d^2}.$$

5.3 Scaling the Data for Numerical Solvers

Strike offsets d and option prices $P(d,t)$ can span large magnitudes, so scaling the data is likely necessary to ensure stable and efficient solver performance.