

Weighted Likelihood Estimation in Quantitative Ecology

Overview of Weighted Likelihood Estimation (WLE)

In the standard maximum likelihood framework, each observation contributes equally to the likelihood. **Weighted Likelihood Estimation (WLE)** modifies this by assigning a **weight or score to each observation**, so that some data points influence the parameter estimates more (or less) than others. In essence, WLE is a *weighted* version of maximum likelihood: observations with higher weights are effectively counted multiple times, whereas those with low weights contribute little to the likelihood (r - How to incorporate weights into a likelihood function? - Stack Overflow) (r - How to incorporate weights into a likelihood function? - Stack Overflow). This approach is useful in quantitative ecology because ecological data often come from heterogeneous sources – for example, surveys with **unequal effort**, varying **detection probabilities**, differing **habitat quality**, or observer-specific reliability. By incorporating external weights, ecologists can adjust the analysis to account for these factors, ensuring that more reliable or representative data have a greater impact on the results than data known to be less reliable.

Why use weights in ecology? Consider a simple example: if a species is not detected during a very intensive survey of a site, that non-detection is more convincing evidence of absence than a non-detection during a cursory survey. WLE provides a formal way to reflect such differences. It bridges the gap between *unweighted* analyses that might ignore important sampling differences and more complex hierarchical models. WLE can be seen as a compromise that integrates extra information (like effort or detection probability) into the likelihood itself, rather than requiring an entirely separate model component for those factors. In practice, WLE has been used both to correct known biases (such as sampling or detection bias) and to achieve robust estimation by down-weighting outliers. The following sections outline the mathematical formulation of WLE and discuss how weights can be derived and applied in ecological contexts.

Standard vs. Weighted Likelihood Framework

Standard MLE: Suppose we have observations x_1, x_2, \dots, x_n and a statistical model with likelihood contributions $L_i(\theta) = f(x_i | \theta)$ for parameter θ . The usual *log-likelihood* is the sum of log-probabilities for each observation:

$$\ell(\theta) = \sum_{i=1}^n \ln f(x_i | \theta).$$

Maximizing $\ell(\theta)$ yields the **maximum likelihood estimate (MLE)** $\hat{\theta}_{MLE}$. This treats each observation equally – each x_i contributes one term to the sum.

Weighted Likelihood: WLE introduces a set of nonnegative weights w_1, w_2, \dots, w_n (provided externally or from prior knowledge) for the observations. The *weighted* log-likelihood is defined as:

$$\ell_w(\theta) = \sum_{i=1}^n w_i \ln f(x_i | \theta).$$

Equivalently, the weighted likelihood can be written as $L_w(\theta) = \prod_{i=1}^n [f(x_i | \theta)]^{w_i}$. Maximizing $\ell_w(\theta)$ gives the **weighted likelihood estimate** $\hat{\theta}_{WLE} = \arg \max_{\theta} \ell_w(\theta)$. The only difference from ordinary MLE is the presence of the weights w_i inside the sum. If all $w_i = 1$, then $\ell_w(\theta)$ reduces to the standard log-likelihood. If a particular $w_j = 0$, observation x_j is effectively *excluded* from the inference. Intermediate values (e.g. $0 < w_i < 1$ or $w_i > 1$) down-weight or up-weight an observation's influence proportionally. In fact, using an integer weight $w_i = m$ is equivalent to treating x_i as m *independent copies* in the dataset (r - How to incorporate weights into a likelihood function? - Stack Overflow). More generally, any real weight can be interpreted as a fractional replication or importance of that data point. For example, an observation with $w_i = 2$ contributes twice as much log-likelihood as one with $w = 1$, and $w = 0.5$ means it contributes half as much as a typical point.

Weighted score equations: The maximization condition for WLE leads to *weighted score equations*. Recall that for MLE, the score equation is $\sum_{i=1}^n u(x_i; \theta) = 0$, where $u(x; \theta) = \partial \ln f(x; \theta) / \partial \theta$ is the score function. In WLE, this generalizes to:

$$\sum_{i=1}^n w_i u(x_i; \hat{\theta}_{WLE}) = 0,$$

meaning each data point's score is scaled by w_i when summing. As a result, $\hat{\theta}_{WLE}$ is the solution where the *weighted average score* is zero. Intuitively, observations with larger weights must have their individual scores closer to zero (or balanced by others) to satisfy this equation, so they pull the estimate more strongly. Observations with tiny weights have little impact on the balance of this equation. Thus, **WLE modifies the influence of each observation on the parameter estimates in a controlled way**, without changing the form of the model's likelihood. All standard properties of likelihood-based estimation (consistency, asymptotic normality, etc.) can still apply under WLE, although one must account for the weighting when deriving variances (discussed below).

Incorporating Weights: Ecological Motivation and Derivation

In ecological studies, weights are typically determined **a priori** based on external information about data quality or sampling design. They are not arbitrary

tuning parameters, but rather reflect known differences in how the data were collected or how reliable each observation is. Below are common scenarios in ecology and how weights can be derived in each case:

- **Unequal Sampling Effort or Selection Probability:** Not all samples are collected with the same effort or probability. For instance, one monitoring site might have been surveyed 10 times, whereas another was surveyed only once; or certain areas of a landscape were sampled more intensively than others. To avoid giving disproportionately high influence to data from highly sampled sites, or conversely to correct under-representation of sparsely sampled areas, we can assign weights based on *sampling effort or design*. One approach is to use **inverse probability weighting**: each observation is weighted by the inverse of the probability that it was included or detected in the sample (modeling - Do ecological studies commonly use weights? - Cross Validated). This ensures that under-sampled units (e.g. remote sites or rare habitat types) get **upweighted** to represent themselves in the population, while over-sampled units are downweighted. In practice, if a site has E_i units of effort (e.g. trap-nights, survey hours), one might set $w_i = E_i$ or w_i proportional to E_i so that sites with more effort contribute more “trials” to the likelihood. Conversely, in a stratified survey where some strata are intentionally over-sampled, weights can standardize contributions to reflect a balanced population view (modeling - Do ecological studies commonly use weights? - Cross Validated). The net effect is that WLE can produce estimates that are **unbiased with respect to the target population** despite unequal sampling. (In fact, treating weights as frequency multipliers is equivalent to the classical Horvitz-Thompson weighting used in survey analysis, but implemented in a likelihood framework.)
- **Imperfect Detection Probability:** A hallmark of ecological data is that failure to observe a species does not guarantee its absence. If the probability of detecting the species (when it is present) is less than 1, standard likelihood analyses that treat nondetection as absence will *bias occupancy or occurrence estimates low*. WLE offers a simplification to incorporate detection probabilities without a fully hierarchical model. In cases where an **external estimate of detection probability** p_i is available for each observation (e.g. from pilot studies or based on conditions), we can weight the observation by that detection probability to adjust its influence. Specifically, **nondetections can be down-weighted** according to the chance that a true presence would have been missed. For example, if a survey at site i has a detection probability $p_i = 0.3$, a failure to detect the species might be given weight $w_i = 0.3$, reflecting that it is only 30% “reliable” evidence of absence. On the other hand, a confirmed detection could be given full weight (or even a higher weight if needed to balance totals) since it definitively indicates presence. The rationale is that an observation with low detectability should not contribute as strongly to likelihood components that assume detection. In effect, this is akin to **scaling the likelihood**

by the detection odds. As an illustration, MacKenzie *et al.* (2002) showed that accounting for detection probability (via a likelihood-based model) raised the estimated occupancy of American toads from 34% (naïve proportion of sites observed) to 49% (ecol_83_824.2248_2255.tp) – a 44% increase – because many sites where the toad was present went undetected in surveys. Ignoring detectability would underestimate occupancy. Weighting can approximate this adjustment: *virtually all abundance or occupancy estimators can be viewed as counts or frequencies divided by detection probability* (ecol_83_824.2248_2255.tp), which conceptually is the same as upweighting the observations by $1/p$ or downweighting incomplete detections by p . In practice, dedicated occupancy models explicitly include p as a parameter; but if one chooses a WLE approach, careful choice of weights derived from p_i can correct much of the bias due to false absences.

- **Habitat Quality or Covariate-Based Weights:** Ecological knowledge about habitat suitability or other covariates can inform the weighting of observations. If a particular site has very high-quality habitat for the species, one *expects* the species to be present; an absence there might be viewed with skepticism (perhaps attributed to unlucky timing or cryptic behavior). WLE allows us to formalize this intuition by giving such a surprising absence a lower weight – effectively saying “don’t let this one observation overly decrease the estimated occurrence, because it might be a false negative.” Conversely, a presence recorded in an extremely poor or atypical habitat might warrant caution (it could be a transient individual or even a misidentification), and one could down-weight that data point so it doesn’t unduly influence habitat preference inference. In both cases, the weight acts as a surrogate for unmodeled factors: high-quality habitat absences are given less influence, and low-quality habitat presences are tempered. These weights might be derived from a habitat suitability index or expert scores. For example, if habitat suitability at a site is quantified on a 0–1 scale, that value or a function of it could serve as w_i . In practice, **direct modeling of habitat quality as a covariate** in the likelihood is usually preferable (Program PRESENCE help file); however, WLE offers a shortcut if one prefers to keep the model simple while still accounting for habitat-based reliability. The key is that weights incorporate our **prior confidence** in the data: we trust absences less in places the species is expected, and perhaps trust presences less in places the species is unexpected. This strategy must be used carefully to avoid introducing subjective bias, but it can improve model fit when certain datapoints are known to be questionable.
- **Observer Reliability and Data Quality:** In many ecological studies (especially citizen-science or multi-observer projects), different observers have different skill levels. Some observers might consistently detect and correctly identify species, while others might miss individuals or make identification errors. *Observer-specific weights* can be assigned to account for this variability in data quality (Understanding the reliability of citizen

science observational data ...). For instance, data contributed by an expert birder could receive a weight $w = 1$ (full trust), whereas data from an inexperienced volunteer might receive $w = 0.5$ to reduce its impact. If an observer is known to have a false-positive rate or false-negative rate, weights might be tied to those metrics (higher weight for more reliable observers). Statistical frameworks for crowdsourced data explicitly use this idea: “*greater weight to information from more competent users*” is given during inference (Understanding the reliability of citizen science observational data ...). A practical example is eBird or other citizen-science bird monitoring programs, where observations are sometimes weighted by an observer’s past performance or by check-list effort. By using WLE, one can integrate these observer weights directly into the likelihood calculation of a species distribution model or occupancy analysis, rather than, say, dropping data or modeling observer as a random effect. This approach ensures that **observer reliability is quantitatively influencing the estimates** – reliable observers essentially count more in determining parameters like occurrence probability or population size, whereas data from potentially unreliable sources contribute less to the final estimate. The result is typically a more robust estimation that acknowledges data quality differences.

In all these cases, the weights w_i are derived from *external considerations* – be it sampling design probabilities, empirical detection rates, habitat-based expectations, or expert judgment. WLE uses these weights to **tilt the likelihood** so that it better reflects the true signal in the presence of known biases or uncertainties. It’s important to note that using WLE does not change the underlying model for the ecological process; it changes how the data inform that model. Essentially, WLE is performing a form of **bias correction or emphasis adjustment** within the likelihood framework.

Theoretical and Practical Considerations

Asymptotic properties and inference: If the weights accurately represent the differing information content or bias of observations, $\hat{\theta}_{WLE}$ can be a **consistent estimator** of the true parameter. For example, in a stratified survey scenario, the weighted likelihood estimator corresponds to a **pseudo-mMaximum likelihood** that targets the population parameters (much like design-based estimators do) (modeling - Do ecological studies commonly use weights? - Cross Validated). However, because the weighted likelihood generally does not correspond to the likelihood of any simple random sample (unless weights are all equal or integer frequencies), the usual formulas for standard errors may not directly apply. In survey analysis, it is common to use a **robust sandwich estimator** of variance to account for the weighting (R: Maximum pseudolikelihood estimation in complex surveys). The sandwich estimator essentially treats the weighted likelihood as an estimating equation and uses the variability of the scores and the information matrix to get correct standard errors under the complex sampling

scheme (R: Maximum pseudolikelihood estimation in complex surveys). If one uses software that supports case weights, it often will automatically compute appropriate standard errors (for example, the R `survey` package or weighted GLM procedures). It's good practice to verify whether a given software's log-likelihood is indeed using w_i as *frequency weights* (replicating data) or as so-called *importance weights* (which might scale variance differently). In most statistical packages and theoretical treatments, case weights are treated as if each observation was replicated w_i times (r - How to incorporate weights into a likelihood function? - Stack Overflow) (Weighted Generalized Linear Models - statsmodels 0.15.0 (+649)), which means the Fisher information is effectively scaled by the weights as well.

Choosing and normalizing weights: The absolute scale of the weights does not affect the location of the maximum likelihood – only their relative ratios matter. If all weights are multiplied by a constant factor, the weighted MLE $\hat{\theta}_{WLE}$ remains the same (since the log-likelihood is just scaled by that constant). However, scaling weights can affect the interpretation of the log-likelihood value and certain information criteria. Some analysts normalize weights to sum to the sample size n or to 1, so that one can still compare likelihood ratios or use AIC on a comparable scale. Others leave weights on their natural scale (e.g., total effort or inverse inclusion probability) and rely on simulation or adjusted criteria for model selection. In ecology, if weights come from probabilities (which sum to n in expectation) or counts of effort, it's often fine to leave them unnormalized (modeling - Do ecological studies commonly use weights? - Cross Validated). If weights are extremely varied, one should be cautious: a very large weight on one data point can overwhelm the likelihood, potentially leading to estimates that mostly reflect that single observation's needs. It is wise to conduct sensitivity analyses – e.g. try capping weights or see how removing high-weight points affects results – to ensure no single weight is unduly driving the inference.

Theoretical robustness vs. model-fit tradeoff: One attractive feature of WLE is its connection to robust statistics. By down-weighting observations that are outlying or suspected to be less reliable, WLE can increase the *robustness* of estimates (reducing the influence of aberrant data). Indeed, the weighted likelihood approach has been shown to produce estimators with high breakdown points and good efficiency. In ecology, this robustness can guard against issues like one faulty sensor that gave an implausible count, or one site that was misclassified as habitat. Instead of excluding such points entirely, WLE offers a nuanced alternative: include them with a low weight. This way, they do not distort the results, but the analysis acknowledges their presence with minimal influence. On the flip side, if weights are mis-specified (e.g., one assumes a detection probability or uses an erroneous effort measure), WLE could introduce bias. Therefore, weights should be chosen based on the best available scientific knowledge or data. They can also be iteratively refined – for example, one might initially weight by effort, fit a model, and then check residuals or leverage points. If some points still appear highly influential or problematic, additional down-weighting (as in robust iterative schemes) could be applied, though this

enters more subjective territory.

Practical examples: To concretize the use of WLE, imagine modeling the occurrence of a bird species across 100 sites, where 50 sites were surveyed twice as long as the other 50. A standard likelihood (say a logistic regression of presence/absence on habitat) would treat all 100 data points equally. But you know the longer surveys are more reliable; to use WLE, you assign $w = 2$ to those observations and $w = 1$ to the shorter surveys. The weighted likelihood then effectively counts each long-survey data point twice. The resulting occurrence estimate will likely be higher than an unweighted analysis if many of the short surveys missed the species (since those misses had weight 1 instead of 2, they contribute less evidence of absence). Another example: combining **presence-only data with survey data**. If you have opportunistic presence records (with no absences recorded) from citizen science and a designed study with known effort and absences, you could assign weights to the opportunistic records to reflect their lower reliability. Perhaps each citizen observation gets $w = 0.5$, while each systematic survey detection gets $w = 1$ (and each nondetection in a systematic survey might get something like $w = 1$ as well or lower if needed). The combined weighted likelihood would then use both data sources, but the citizen science data (which might be numerous but less controlled) won't overwhelm the inference – their influence is tempered by the weight. This kind of weighting strategy has been recommended for integrating **crowd-sourced data** with traditional data (Understanding the reliability of citizen science observational data ...).

In summary, **Weighted Likelihood Estimation** is a powerful extension of the likelihood paradigm that introduces flexibility for handling common issues in ecological data. By incorporating external weights, analysts can adjust for unequal sampling effort, correct biases from imperfect detection, and account for known differences in data quality or relevance. The mathematical foundation of WLE is straightforward – it's the same likelihood equations with a weighted sum – but the practical effect is a model that is **informed by more than just the raw data**. It leverages prior knowledge and study design information to produce estimates that are often more realistic and less biased. When using WLE, ecologists should justify their choice of weights and be mindful of the implications for uncertainty estimates. With appropriate precautions, WLE serves as a valuable tool in the toolkit of ecological statistics, allowing more nuanced and credible modeling of complex natural systems.

Sources:

- Julius Vainora (StackOverflow answer) – Demonstration of implementing weighted log-likelihood (case weights) (r - How to incorporate weights into a likelihood function? - Stack Overflow) (r - How to incorporate weights into a likelihood function? - Stack Overflow).
- Cross-Validated (AdamO) – Discussion on when to use sampling weights in ecological and public health studies (modeling - Do ecological studies commonly use weights? - Cross Validated).

- MacKenzie *et al.* (2002) *Ecology* 83:2248 – Occupancy modeling accounting for detection < 1 (detection bias leads to underestimation of occupancy) (ecol_83_824.2248_2255.tp) (ecol_83_824.2248_2255.tp).
- Statsmodels documentation – Interpretation of frequency weights as replicating observations (Weighted Generalized Linear Models - statsmodels 0.15.0 (+649)).
- Bird *et al.* (2014) and Johnston *et al.* (in prep) via BES methods – Notion of weighting more competent observers in citizen science (Understanding the reliability of citizen science observational data ...).
- Agostinelli (2001) – Robust statistics via weighted likelihood (WLE yields high-breakdown robust estimators).
- Lumley (R **survey** package) – Pseudo-likelihood and robust variance for complex survey data (R: Maximum pseudolikelihood estimation in complex surveys).