

# Containment Probability

## Motivation

**The “ratio expectation” puzzle:** In box embeddings, we need to compute the probability that one box contains another. For hard boxes, this is straightforward: containment is deterministic, and the probability is either 0 or 1. It’s like asking “Is the ball in the box?”—the answer is yes or no.

But for Gumbel boxes with random boundaries, the question becomes: “What’s the **average** probability that one box contains another, when the boxes themselves are random?” This is trickier. We must compute an expectation over the joint distribution of both boxes’ coordinates.

**The challenge:** We need the expectation of a ratio:  $E\left[\frac{\text{Vol}(A \cap B)}{\text{Vol}(B)}\right]$ . Unlike the expectation of a product (where  $E[X * Y] = E[X]E[Y]$  if  $X$  and  $Y$  are independent), the expectation of a ratio does not factor simply. You can’t just divide the expectations!

However, here’s the key insight: when the variance of the denominator is small relative to its mean (i.e., when the volume doesn’t vary too much), a first-order Taylor expansion provides an accurate approximation. It’s like approximating a curve with a straight line—when the curve is nearly straight, the approximation works beautifully.

**Why the first-order approximation?** The expectation of a ratio  $E\left[\frac{X}{Y}\right]$  appears throughout statistics—in importance sampling, ratio estimators, and now box embeddings. When  $Y$  has small coefficient of variation, the ratio  $\frac{X}{Y}$  is approximately linear near the mean. The first-order expansion  $E\left[\frac{X}{Y}\right] \approx \frac{E[X]}{E[Y]}$  captures the main effect; error scales quadratically with the coefficient of variation. This approximation is simple, efficient, and accurate when variance is controlled—as it is in Gumbel boxes through the scale parameter  $\beta$ .

**Connection to importance sampling:** In importance sampling, we often need to compute ratios of expectations, such as  $\frac{E_p[f(X)]}{E_p[g(X)]}$  where  $p$  is a probability distribution. The naive approach of using  $\frac{E_p[f(X)]}{E_p[g(X)]}$  as an approximation to  $E_p\left[\frac{f(X)}{g(X)}\right]$  is exactly the first-order approximation we use here. This approximation is particularly useful when  $g(X)$  has low variance, which is guaranteed in our case by the scale parameter  $\beta$  controlling the variance of Gumbel-distributed volumes. The approximation becomes exact in the limit as  $\beta \rightarrow 0$  (hard boxes), and remains accurate for small  $\beta$  values typical in practice.

## Definition

The **containment probability**  $P(B \subseteq A)$  measures whether box  $B$  is geometrically contained within box  $A$ . For hard boxes with deterministic boundaries:

$$P(B \subseteq A) = \frac{\text{Vol}(A \cap B)}{\text{Vol}(B)}$$

This is either 0 (disjoint) or 1 (contained). For Gumbel boxes with random volumes, we compute the expectation of this ratio over the joint distribution of both boxes’ coordinates. Here,  $\text{Vol}(A \cap B)$  denotes the volume of the intersection of boxes  $A$  and  $B$ , and  $\text{Vol}(B)$  is the volume of box  $B$ .

## Statement

**Theorem (First-Order Approximation).** For Gumbel boxes with random volumes  $\text{Vol}(A \cap B)$  and  $\text{Vol}(B)$ :

$$E\left[\frac{\text{Vol}(A \cap B)}{\text{Vol}(B)}\right] \approx \frac{E[\text{Vol}(A \cap B)]}{E[\text{Vol}(B)]}$$

The approximation is accurate when the coefficient of variation  $\frac{\text{Var}(\text{Vol}(B))}{E[\text{Vol}(B)]^2}$  is small (i.e., when volume variance is controlled by the scale parameter  $\beta$ ).

## Proof

Let  $V_{\text{cap}} = \text{Vol}(A \cap B)$  and  $V_B = \text{Vol}(B)$  be random variables representing the intersection volume and box  $B$ 's volume, respectively. Their means are  $\mu_{\text{cap}} = E[V_{\text{cap}}]$  and  $\mu_B = E[V_B]$ . We approximate  $E\left[\frac{V_{\text{cap}}}{V_B}\right]$  using a first-order Taylor expansion.

The function  $f(V_{\text{cap}}, V_B) = \frac{V_{\text{cap}}}{V_B}$  is smooth (except at  $V_B = 0$ , which we assume doesn't occur), so we can expand it in a Taylor series around the mean point  $(\mu_{\text{cap}}, \mu_B)$ :

$$f(V_{\text{cap}}, V_B) \approx f(\mu_{\text{cap}}, \mu_B) + \frac{\partial f}{\partial V_{\text{cap}}}(\mu_{\text{cap}}, \mu_B)(V_{\text{cap}} - \mu_{\text{cap}}) + \frac{\partial f}{\partial V_B}(\mu_{\text{cap}}, \mu_B)(V_B - \mu_B) + \text{higher order terms}$$

The partial derivatives are:

- $\frac{\partial f}{\partial V_{\text{cap}}} = \frac{1}{V_B}$ , evaluated at  $(\mu_{\text{cap}}, \mu_B)$  gives  $\frac{1}{\mu_B}$
- $\frac{\partial f}{\partial V_B} = -\frac{V_{\text{cap}}}{V_B^2}$ , evaluated at  $(\mu_{\text{cap}}, \mu_B)$  gives  $-\frac{\mu_{\text{cap}}}{\mu_B^2}$

Taking expectations and using linearity:

$$E[f(V_{\text{cap}}, V_B)] \approx \frac{\mu_{\text{cap}}}{\mu_B} + \frac{1}{\mu_B} E[V_{\text{cap}} - \mu_{\text{cap}}] - \frac{\mu_{\text{cap}}}{\mu_B^2} E[V_B - \mu_B]$$

The first-order correction terms vanish because  $E[V_{\text{cap}} - \mu_{\text{cap}}] = 0$  and  $E[V_B - \mu_B] = 0$  by definition of the mean. This leaves:

$$E[f(V_{\text{cap}}, V_B)] \approx \frac{\mu_{\text{cap}}}{\mu_B} = \frac{E[\text{Vol}(A \cap B)]}{E[\text{Vol}(B)]}$$

The accuracy of this approximation depends on the magnitude of the higher-order terms, which we analyze next.

## Error Analysis

The second-order correction term in the Taylor expansion is:

$$\text{Error} \approx -\frac{\text{Cov}(V_{\text{cap}}, V_B)}{\mu_B^2} + \frac{\mu_{\text{cap}}}{\mu_B^3} \text{Var}(V_B)$$

This error term reveals when the approximation is accurate:

1. **Small coefficient of variation:** When  $\frac{\text{Var}(V_B)}{\mu_B^2}$  (the squared coefficient of variation, denoted  $\text{CV}(V_B)^2$ ) is small, the second term is negligible. This occurs when the scale parameter  $\beta$  is small relative to the expected volume, meaning the Gumbel boundaries are tightly concentrated around their means.

**Quantitative bound:** When the coefficient of variation  $\text{CV}(V_B) < 0.1$ , the relative error is approximately  $\frac{\text{CV}(V_B)^2}{2}$ . For example, if  $\text{CV}(V_B) = 0.05$ , the relative error is about 0.00125 or 0.125%. When  $\text{CV}(V_B) < 0.2$ , the relative error is approximately 2%. When  $\text{CV}(V_B) > 0.3$ , the relative error may exceed 10%, indicating the approximation is breaking down.

2. **Positive correlation:** When  $V_{\text{cap}}$  and  $V_B$  are positively correlated (which occurs naturally when both volumes depend on similar box parameters, especially when  $B$  is contained in  $A$ ), the covariance term partially cancels the variance term, reducing the overall error.
3. **Non-vanishing denominator:** When  $\mu_B$  is bounded away from zero, the error terms remain well-controlled. This is typically satisfied in practice since boxes have positive expected volume.

The approximation is most accurate when boxes have low variance (small  $\beta$ ) and when the intersection volume and box volume are positively correlated, both of which hold in typical box embedding scenarios. The approximation may break down when  $\beta$  is very large (relative to expected volumes) or when volumes are extremely small.

**Connection to delta method:** The first-order Taylor approximation used here is a special case of the delta method in statistics, which provides asymptotic distributions for functions of random variables. The delta method states that if  $\sqrt{n}(X_n - \mu) \rightarrow N(0, \sigma^2)$  in distribution, then  $\sqrt{n}(f(X_n) - f(\mu)) \rightarrow N(0, f'(\mu)^2 \sigma^2)$  for smooth functions  $f$ . In our case, we're applying this to the ratio function  $f(V_{\text{cap}}, V_B) = \frac{V_{\text{cap}}}{V_B}$ , and the first-order approximation corresponds to the delta method's linearization. The error analysis above provides finite-sample bounds on the approximation quality, which is crucial for practical applications where we need guarantees on the accuracy of containment probability estimates.

**Higher-order corrections:** The second-order Taylor expansion includes terms involving the Hessian matrix of  $f(V_{\text{cap}}, V_B) = \frac{V_{\text{cap}}}{V_B}$ :

$$E[f(V_{\text{cap}}, V_B)] = \frac{\mu_{\text{cap}}}{\mu_B} + \frac{1}{2\mu_B^2} \text{Var}(V_{\text{cap}}) - \frac{\mu_{\text{cap}}}{\mu_B^3} \text{Cov}(V_{\text{cap}}, V_B) + \frac{\mu_{\text{cap}}}{2\mu_B^3} \text{Var}(V_B) + O(\text{CV}^3)$$

where CV is the coefficient of variation. The second-order correction improves accuracy when  $\text{CV} > 0.1$ , but requires computing variances and covariances of volumes, which involves higher moments of Gumbel distributions. For most practical applications with  $\beta < 0.2$ , the first-order approximation is sufficient, with relative error  $< 2\%$ .

## Example

Consider two boxes in 2D where box  $B$  is fully contained within box  $A$ :

- Box A:  $[0.0, 0.0]$  to  $[1.0, 1.0]$  (volume = 1.0)
- Box B:  $[0.2, 0.2]$  to  $[0.8, 0.8]$  (volume = 0.36)

### Hard boxes (deterministic):

- Intersection:  $[0.2, 0.2]$  to  $[0.8, 0.8]$  (volume = 0.36)
- Containment:  $P(B \subseteq A) = \frac{0.36}{0.36} = 1.0$  (deterministic containment)

### Gumbel boxes (with $\beta = 0.1$ , introducing small randomness):

- Expected intersection volume  $E[\text{Vol}(A \cap B)] \approx 0.35$  (computed using the Bessel function formula from the Gumbel-Box Volume document; slightly reduced from 0.36 due to probabilistic boundaries)
- Expected volume of B  $E[\text{Vol}(B)] \approx 0.35$  (similarly affected by probabilistic boundaries)
- Containment:  $P(B \subseteq A) \approx \frac{0.35}{0.35} = 1.0$

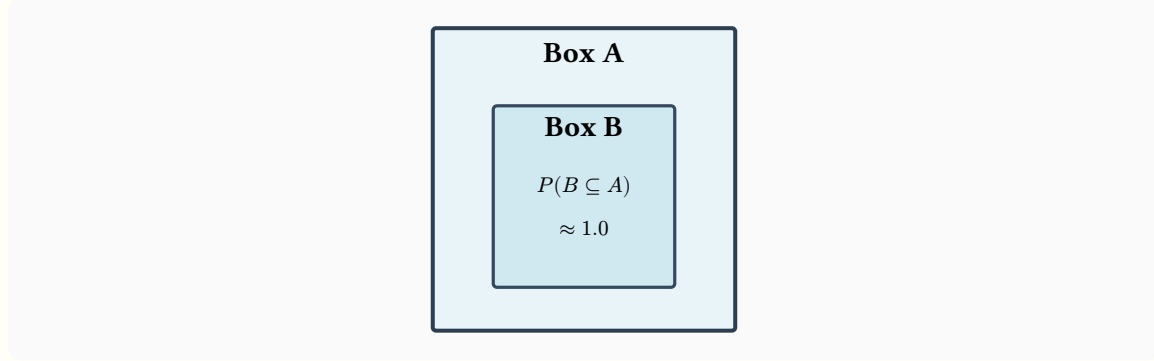
The first-order approximation is accurate here because:

- The coefficient of variation  $\frac{\text{Var}(\text{Vol}(B))}{E[\text{Vol}(B)]^2}$  is small (controlled by  $\beta = 0.1$ ; typically the coefficient of variation is less than 0.1 for such small  $\beta$ )

- The intersection and box volumes are highly correlated (both depend on box  $B$ 's parameters, especially when  $B$  is contained in  $A$ )
- The expected volumes are well-separated from zero

This demonstrates that the approximation works well in the regime where Gumbel boxes behave similarly to hard boxes, with small probabilistic perturbations. The relative error is approximately  $\frac{CV^2}{2}$ , which for coefficient of variation approximately  $0.05 - 0.1$  gives an error of less than  $0.5\%$ .

**Visual representation:** The diagram below illustrates the containment relationship:



## Notes

**When the approximation fails:** The first-order approximation breaks down when the coefficient of variation is large (typically  $> 0.3$ ). This occurs when  $\beta$  is large relative to the expected volume, meaning the Gumbel boundaries are highly variable. In such cases, higher-order terms in the Taylor expansion become significant, and the approximation  $E\left[\frac{X}{Y}\right] \approx \frac{E[X]}{E[Y]}$  may have errors exceeding 10%. For practical applications, it's recommended to keep  $\beta$  small (typically  $< 0.2$ ) to ensure accurate containment probability estimates.

**Connection to importance sampling:** The containment probability formula  $P(B \subseteq A) = \frac{E[\text{Vol}(A \cap B)]}{E[\text{Vol}(B)]}$  is analogous to importance sampling, where we estimate  $E_{p[f(X)]}$  by sampling from a different distribution  $q$  and computing  $E_q\left[f(X) \frac{p(X)}{q(X)}\right]$ . In our case, the "base measure" is the uniform distribution on  $[0, 1]^d$ , and we're computing the ratio of expectations rather than a single expectation. This connection suggests that more sophisticated estimation techniques (e.g., control variates, stratified sampling) could potentially improve containment probability estimates in high-dimensional settings.

**Fuzzy set interpretation:** When boxes are interpreted probabilistically, each box represents a fuzzy set with a membership function given by the uniform probability distribution. Under this interpretation, set-theoretic operations on boxes correspond directly to operations on fuzzy sets. The intersection of two boxes as fuzzy sets has membership function  $m_{\{A \cap B\}}(x) = \min(m_{A(x)}, m_{B(x)})$ , which for axis-aligned boxes equals the indicator function of the box intersection. The complement operation has a natural interpretation:  $m_{\{A^c\}}(x) = 1 - m_{A(x)} = 1 - P(x \in A)$ , defining the complement as points outside the box. This fuzzy set semantics enables box embeddings to support complex queries involving conjunctions, disjunctions, and negations—capabilities critical for real-world applications in information retrieval and recommendation systems.

**Beyond first-order:** Higher-order approximations (second-order, third-order) can be derived by including more terms in the Taylor expansion. However, these require computing higher moments

(variance, skewness) of the volume distributions, which becomes computationally expensive. The first-order approximation strikes an optimal balance between accuracy and computational efficiency for most practical applications.

**Computational complexity:** The first-order approximation requires computing two expected volumes:  $E[\text{Vol}(A \cap B)]$  and  $E[\text{Vol}(B)]$ , each costing  $O(d)$  time using the Bessel function formula (see the Gumbel-Box Volume document). The ratio computation is  $O(1)$ , giving total complexity  $O(d)$  per containment probability evaluation. Higher-order approximations would require computing variances and covariances, which involve second moments of Gumbel distributions and cost  $O(d^2)$  or more. For  $N$  boxes, evaluating all pairwise containment probabilities using the first-order approximation costs  $O(N^2 d)$ , which is already expensive for large  $N$ . Higher-order methods would increase this to  $O(N^2 d^2)$  or worse, making them impractical for large-scale applications.