Coverage-at-K/Q: Simple and Interpretable Sharpness Metrics

Keunwoo Choi[1] and Kyunghyun Cho[*2]

[1]KAIST
[2]New York University

Abstract

We introduce two lightweight and interpretable metrics for describing the sharpness (or conversely, the spread) of a categorical distribution: Coverage-at-K ($C(K)$) and Coverage-at-Q ($\overline{C}(Q)$). $C(K)$ operates on raw counts and is summarized by a normalized area-under-the-curve metric, AUC-C. $\overline{C}(Q)$ operates on probabilities and is summarized by a Uniform Divergence Score (UDS) score. Both metrics are derived by relaxing the binary notion of coverage (i.e., the fraction of non-empty categories) by varying a threshold and analyzing the resulting curve. This approach provides a richer, more robust characterization of distributional sharpness than simple coverage or entropy alone.

1 Motivation

Measuring the uncertainty or concentration of a categorical distribution is a fundamental task in many fields. The most common metric is Shannon entropy [1], which quantifies the average level of "information" or "surprise" inherent in a variable's possible outcomes. While mathematically rigorous, the value of entropy (in bits or nats) can be unintuitive to non-specialists. Other measures of statistical dispersion, such as the Gini coefficient [2] from economics or Simpson's index [3] from ecology, quantify inequality but can also be difficult to interpret directly in terms of category coverage.

A simpler alternative is plain coverage: the fraction of categories that are observed (i.e., have a non-zero count). However, this metric is brittle, as a single spurious observation can inflate its value, making a highly skewed distribution appear uniform.

To address these shortcomings, we propose metrics based on a thresholding approach, analogous to the Receiver Operating Characteristic (ROC) curve analysis widely used in machine learning [4]. Instead of a single value, we analyze how coverage decays as we increase the minimum acceptable count (K) or probability (Q) for a category to be considered "covered." This yields two related families of metrics: Coverage-at-K ($C(K)$) for counts and Coverage-at-Q ($\overline{C}(Q)$) for probabilities.

2 Coverage-at-K ($C(K)$) and AUC-C

2.1 Coverage-at-K ($C(K)$)

Given a multiset of observations represented by a vector of counts over a universe of C_{total} possible categories, we define Coverage-at-K as the fraction of categories whose count exceeds a threshold K.

*Corresponding author: kyunghyun.cho@nyu.edu

Definition. Let $\mathcal{C}$ be the set of all possible categories. For a given count function $\mathrm{count}(c)$ for each category $c \in \mathcal{C}$, Coverage-at-K is:

$$C(K) = \frac{1}{|\mathcal{C}|}|\{c \in \mathcal{C} \mid \mathrm{count}(c) > K\}|$$

The strict inequality ($> K$) ensures that C(0) corresponds exactly to the ordinary definition of coverage (the proportion of non-empty categories). The function $C(K)$ is non-increasing in K and eventually reaches 0 for $K \geq \max_{c\in\mathcal{C}}(\mathrm{count}(c))$.

2.2 Even Point and AUC-C

To summarize the $C(K)$ curve into a single scalar, we compute its area up to a meaningful reference point. We define this reference as the *even point*, which is the count each observed category would have if the total items were distributed perfectly evenly among them.

Let $N = \sum_c \mathrm{count}(c)$ be the total number of items. The even point is defined as:

$$K_{\mathrm{even}} = \lfloor N/|\mathcal{C}| \rfloor$$

We define the Area Under the Curve for $C(K)$ (AUC-C) by summing $C(K)$ from $K = 0$ to $K_{\mathrm{even}} - 1$ and dividing it by the area of an ideal curve. An ideal distribution (uniform among all categories) would have a $C(K)$ value of 1 for all $K < K_{\mathrm{even}}$.

Definition. The AUC-C is the ratio of the observed area to the ideal area:

$$\text{AUC-}C = \frac{1}{K_{\mathrm{even}}} \sum_{K=0}^{K_{\mathrm{even}}-1} C(K)$$

This metric is bounded in $[0, 1]$. A value of 1.0 implies that the observations are as evenly distributed among all the categories as possible. Values closer to 0 indicate that the mass is concentrated in a few categories, and the minimum is at $\frac{1}{|\mathcal{C}|}$.

3 Coverage-at-Q ($\overline{C}(Q)$) and Uniform Divergence Score (UDS)

3.1 Coverage-at-Q ($\overline{C}(Q)$)

A parallel metric can be defined for probability distributions. Given a probability vector $\mathbf{p} = (p_1, p_2, \ldots, p_C)$ where $\sum_{c=1}^{C} p_c = 1$ and $p_c \geq 0$ for all c, we define Coverage-at-Q as the fraction of categories whose probability is at least Q.

Definition. For a probability threshold $Q \in [0, 1)$, Coverage-at-Q is:[1]

$$\overline{C}(Q) = \frac{1}{|\mathcal{C}|}|\{c \mid p_c > Q\}|$$

$\overline{C}(Q; \mathbf{p})$ is a non-increasing step function and vanishes to 0 when $Q \geq \max_{c\in\mathcal{C}} p_c$, similarly to $C(K)$ above. For a uniform distribution, where $p_c = 1/C$ for all c, $\overline{C}(Q)$ is a simple step function: it is 1 for $Q < 1/C$ and 0 for $Q \geq 1/C$.

3.2 Uniform Convergence Score (UCS)

We can summarize the $\overline{C}(Q)$ curve by measuring its total similarity to the step function of a uniform distribution. Let $q^* = 1/C$, similar to how we defined K_{even} earlier. The Uniform Convergence Score (UCS) measures how similar the $\overline{C}(Q)$ curve is to the uniform reference curve, normalized normalized to lie in $[0, 1]$.

[1] We define $\overline{C}(1) = \lim_{Q\to 1} \overline{C}(Q) = 1$ for the convenience.

Definition. The UCS for a probability vector $\mathbf{p}$ is:

$$\text{UCS}(\mathbf{p}) = 1 - \frac{|\mathcal{C}|^2}{2(|\mathcal{C}|-1)} \left[\int_0^{q^*} (1 - \overline{C}(Q;\mathbf{p}))\, dQ + \int_{q^*}^1 \overline{C}(Q;\mathbf{p})\, dQ \right],$$

where we used $\overline{C}(Q;\mathbf{p})$ to specify that it was computed against the probability vector $\mathbf{p}$.

The UCS is 1 for a uniform distribution and approaches 0 for a maximally sharp (one-hot) distribution. In practice, the integral is computed as a sum, leveraging the fact that the $\overline{C}(Q)$ curve only changes at values Q equal to the probabilities p_c.

4 Noteworthy Properties

In this section, we explore properties of the proposed metriecs, $C(K)$, AUC-C, $\overline{C}(Q)$ and UCS, among each other as well as against other more established metrics. These properties enable us to understand structures behind these metrics better and can guide us how we can choose and use these metrics in practice.

4.1 Formal Relationship between $C(K)$ and $\overline{C}(Q)$

First, we consider the relationship between $C(K)$ and $\overline{C}(Q)$, although they operate on different domains (counts vs. probabilities). The function $\overline{C}(Q)$ can be viewed as a version of the count-based coverage curve where the threshold axis is rescaled by the total number of items, N. This relationship, presented below, enables us to transform Q into an appropriate K without having to normalize the count vector.

Proposition 1. *The relationship between Coverage-at-K and Coverage-at-Q can be stated by the relationship between the total item count N, the count threshold K and the probability threshold Q.*

Let a distribution be defined by a count function count(c) over $|\mathcal{C}|$ categories, with a total of $N = \sum_c$ count(c) items. For any probability threshold $Q \in [0,1]$, the value of $\overline{C}(Q)$ is exactly equal to the value of $C(K)$ for $K = \lfloor N \cdot Q \rfloor$.

$$\overline{C}(Q) = C(\lfloor N \cdot Q \rfloor)$$

Proof. We present the proof in Appendix A. □

4.2 Inverse Monotonic Relationship between AUC-C/UCS and the Sharpness of a Distribution

We show how the proposed metrics, AUC-C and UCS, derived from $C(K)$ and $\overline{C}(Q)$, respectively are related to the uncertainty or sharpness of a distribution. More specifically, we demonstrate that both of these metrics have an inverse monotonic relationship with the sharpness of the distribution. As a distribution becomes more uniform (less sharp), both AUC-C and UCS will increase (or stay the same).

Theorem 2. *The Area Under the Curve for $C(K)$ (AUC-C) is a* ***Schur-concave*** *function of the count vector. This implies that if the count vector $\mathbf{count}_p$ corresponding to a distribution p majorizes the count vector $\mathbf{count}_q$ for a distribution q ($\mathbf{count}_p \succ \mathbf{count}_q$), then:*

$$AUC\text{-}C(\mathbf{count}_p) \leq AUC\text{-}C(\mathbf{count}_q)$$

Proof. We present the proof in Appendix B. □

Theorem 3. *The Uniform Convergence Score (UCS) is a* ***Schur-concave*** *function.*

This implies that if a probability distribution p majorizes a distribution q ($p \succ q$), then:

$$UCS(p) \leq UCS(q)$$

Proof. We present the proof in Appendix C. □

This importan property is shared with a more established measure of the sharpness, or uncertainty, of a distribution, such as the Shannon entropy. The proposed metrics is however less sensitive to how the probability mass is distributed among the observed categories than these conventional measures, which is a feature that makes it more interpretable.

4.3 AUC-C, UCS, Total Variation and Shannon Entropy

Although both AUC-C and UCS were derived from the shared principle of counting non-zero categories after thresholding, either at a certain count or at a certain probability, they reflect two different ways to measure the uncertainty of a distribution. Between these two, AUC-C is more unique in that its sensitivity to the changes in the target count vector is "capped," as the count is flattened out above the even point K_{even}. In other words, AUC-C can be useful for applications where we are interested in the changes in low-count categories but are disinterested once the count of each categories goes above the even point.

On the other hand, UCS is sensitive to the changes in the probability of any category at the whole range $[0, 1]$. We should however characterize this sensitivity, as different measures tend to exhibit distinct types of sensitivity. For instance, the Shannon entropy is significantly more sensitive to smaller probabilities than to larger probabilities, due to the inclusion of log in its definition:

$$\mathcal{H}(\mathbf{p}) = -\sum_{c \in C} p_c \log p_c \in [0, 1]. \tag{1}$$

Unlike the Shannon entropy, UCS's sensitivity is uniform over $[0, 1]$. This uniformity becomes obvious, as we make the following (surprising) realization that UCS, derived from $\overline{C}(Q)$ that measures the thresholded coverage, is a linear function of total variation distance (TVD)[2] against the uniform distribution. [7]

Proposition 4 (Linear Relationship between UCS and TVD to Uniform)**.** *Let $\mathbf{p} = (p_1, \ldots, p_{|\mathcal{C}|})$ be a probability distribution over a set of $|\mathcal{C}|$ categories, and let $\mathbf{u}$ be the uniform distribution over the same categories, where $u_i = 1/|\mathcal{C}|$ for all i.*

The Uniform Convergence Score (UCS) of $\mathbf{p}$ is a direct linear transformation of the Total Variation Distance (TVD) between $\mathbf{p}$ and $\mathbf{u}$. Their relationship is given by the exact formula:

$$UCS(\mathbf{p}) = 1 - \frac{|\mathcal{C}|}{|\mathcal{C}| - 1} \cdot TVD(\mathbf{p}, \mathbf{u})$$

Proof. We present the proof in Appendix D. □

These properties and connections suggest we must choose an appropriate measure and also how to present it according to target applications. In particular, for applications where the focus in on the rare categories and their changes, AUC-C is appropriate. When we know the level of noise in observation, on the other hand, it may be more adequate to report $C(K)$ and/or $\overline{C}(Q)$ directly, over a few possible K/Q threshold values.

[2] TVD is defined as

$$\text{TVD}(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i|. \tag{2}$$

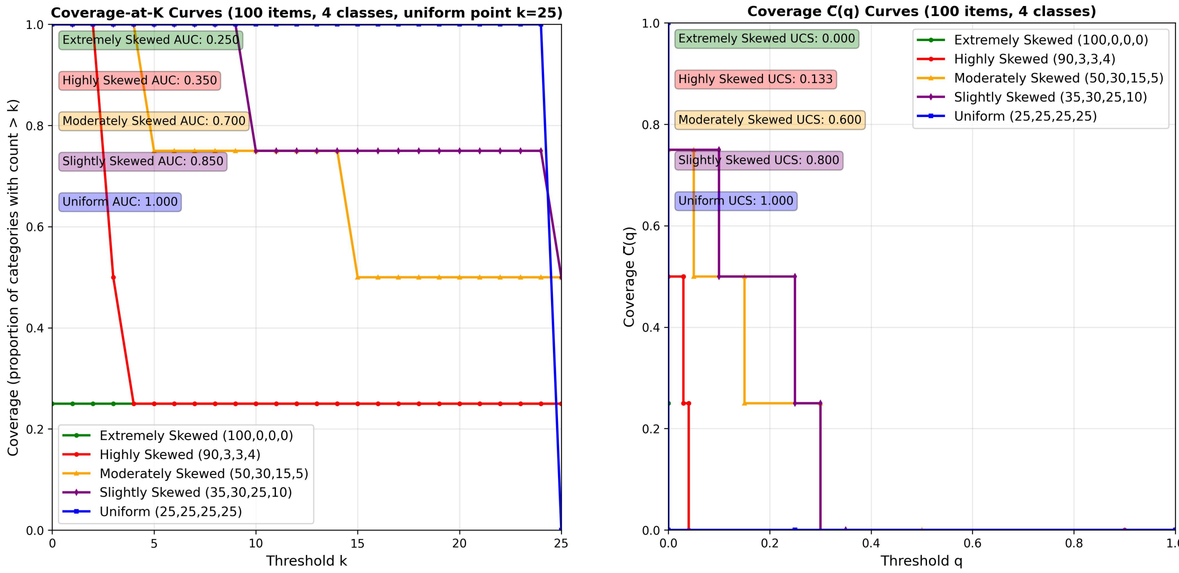

Figure 1: Coverage-at-K (left) and Coverage-at-Q (right) curves for distributions with different levels of skew. Sharper distributions lead to faster decay in the curves.

5 Illustrative Examples

Figure 1 illustrates the $C(K)$ and $\overline{C}(Q)$ curves for four distributions with varying levels of skew. Table 1 shows the corresponding metric values as well as the Shannon Entropy. As skewness increases, AUC-C and UCS both decrease from 1 toward 0. Notably, C(0) as well as $\overline{C}(0)$ both fail to distinguish between any of these cases, except for the degenerate case, highlighting the value of our proposed metrics.

Table 1: Metric values for four distributions (100 items, 4 categories).

Distribution (Counts)	$C(0)$	$\overline{C}(0)$	AUC-C	UCS	Shannon Entropy
Uniform (25, 25, 25, 25)	1.0	1.0	1.000	1.000	2
Slightly skewed (35, 30, 25, 10)	1.0	1.0	0.850	0.800	1.883
Moderately skewed (50, 30, 15, 5)	1.0	1.0	0.700	0.600	1.648
Highly skewed (90, 3, 3, 4)	1.0	1.0	0.350	0.133	0.626
Extremely skewed (100, 0, 0, 0)	0.25	0.25	0.250	0.000	0

In order to understand the relationship among these three measures, AUC-C, UCS and the Shannon entropy, we draw 5,000 10-dimensional probability vectors from Drichlet distribution, compute these measures on each of these vectors and plot three scatter plots illustrating pairwise differences, in Figure 2. We notice a clear linear relationship between AUC-C and UCS, although we notice this relationship is noisy as the distribution approaches uniformity. The relationship between each of the proposed summary statistics and the Shannon entropy is highly non-trivial, and these two plots (middle and right) also highlight how AUC-C and UCS differ from each other.

6 Conclusion

We have presented $C(K)$ and $\overline{C}(Q)$, two simple and interpretable metrics for characterizing the sharpness of categorical distributions. By analyzing coverage as a function of a threshold, they

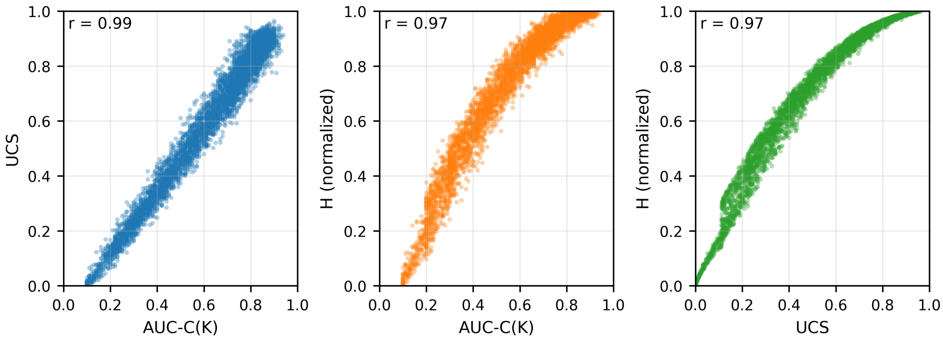

Figure 2: Pairwise scatter plots demonstrating the relationship among the proposed summary statistics, AUC-C, UCS and Shannon entropy. According to these plots, AUC-C and UCS are closely related to each other, while the Shannon entropy shows a nonlinear relationship with both of these measures.

provide a more nuanced view than traditional single-value metrics like entropy or raw coverage. Their summary statistics, AUC-C and UCS, offer robust scalar values to compare distributions, making them useful tools for exploratory data analysis and model evaluation.

Code Availability

You can find the complete implementation of the proposed metrics in Python at https://github.com/keunwoochoi/coverage-at-k.

Declaration on AI Assistants

The authors used Google Gemini 2.5 Pro for drafting this note and both Google Gemini 2.5 Pro and OpenAI GPT-5 for implementing the proposed metrics.

References

[1] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October, 1948.

[2] C. Gini, "Variabilità e mutabilità," Reprinted in *Memorie di metodologica statistica* (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi, 1955.

[3] E. H. Simpson, "Measurement of diversity," *Nature*, vol. 163, p. 688, 1949.

[4] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[5] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer, 2nd edition, 2011.

[6] M. O. Lorenz. "Methods of measuring the concentration of wealth." *Publications of the American Statistical Association*, 9(70):209–219, 1905.

[7] Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). John Wiley & Sons.

A Proof of Proposition 1: Formal Relationship between $\overline{C}(Q)$ and $C(K)$

Proposition 1. *The relationship between Coverage-at-K and Coverage-at-Q can be stated by the relationship between the total item count N, the count threshold K and the probability threshold Q.*

Let a distribution be defined by a count function count(c) over $|\mathcal{C}|$ categories, with a total of $N = \sum_c count(c)$ items. For any probability threshold $Q \in [0, 1]$, the value of $\overline{C}(Q)$ is exactly equal to the value of $C(K)$ for $K = \lfloor N \cdot Q \rfloor$.

$$\overline{C}(Q) = C(\lfloor N \cdot Q \rfloor)$$

Proof. By definition, Coverage-at-Q, $\overline{C}(Q)$, is the fraction of categories whose probability p_c is strictly greater than Q:

$$\overline{C}(Q) = \frac{1}{|\mathcal{C}|}|\{c \in \mathcal{C} \mid p_c > Q\}|$$

And Coverage-at-K, $C(K)$, is the fraction of categories whose count is strictly greater than K:

$$C(K) = \frac{1}{|\mathcal{C}|}|\{c \in \mathcal{C} \mid \text{count}(c) > K\}|$$

To prove the proposition, we must show that the set of categories satisfying the condition for $\overline{C}(Q)$ is identical to the set of categories satisfying the condition for $C(K)$ when we set $K = \lfloor N \cdot Q \rfloor$. This requires proving the following logical equivalence for any category c:

$$p_c > Q \iff \text{count}(c) > \lfloor N \cdot Q \rfloor$$

We begin with the left-hand side (LHS) of the equivalence. Substituting the definition $p_c = \frac{\text{count}(c)}{N}$, we get:

$$\frac{\text{count}(c)}{N} > Q$$

Since the total count N is a positive integer, we can multiply both sides by N without changing the direction of the inequality:

$$\text{count}(c) > N \cdot Q$$

Now, we must prove that for an integer count $\text{count}(c)$, the condition $\text{count}(c) > N \cdot Q$ is equivalent to $\text{count}(c) > \lfloor N \cdot Q \rfloor$. Let $m = \text{count}(c)$ be an integer and $x = N \cdot Q$ be a real number. The equivalence to prove is:

$$m > x \iff m > \lfloor x \rfloor$$

We prove this in two parts.
($\implies$) Assume $m > x$. By the definition of the floor function, $x \geq \lfloor x \rfloor$. From these two inequalities, we have $m > x \geq \lfloor x \rfloor$, which by transitivity implies $m > \lfloor x \rfloor$.
($\impliedby$) Assume $m > \lfloor x \rfloor$. Since both m and $\lfloor x \rfloor$ are integers, this is equivalent to the condition $m \geq \lfloor x \rfloor + 1$. By the definition of the floor function, we also know that for any real number x, $\lfloor x \rfloor + 1 > x$. Combining these inequalities, we have $m \geq \lfloor x \rfloor + 1 > x$, which implies $m > x$.

Since both directions of the implication hold, the conditions are logically equivalent. Therefore, the sets are identical:

$$\{c \in \mathcal{C} \mid p_c > Q\} = \{c \in \mathcal{C} \mid \text{count}(c) > \lfloor N \cdot Q \rfloor\}$$

As the sets are the same, their cardinalities are equal, and thus the proposition $\overline{C}(Q) = C(\lfloor N \cdot Q \rfloor)$ is true. □

B Proof of Theorem 2: AUC-C is a Schur-concave function

Theorem 2. *The Area Under the Curve for $C(K)$ (AUC-C) is a **Schur-concave** function of the count vector. This implies that if the count vector $\mathbf{count}_p$ corresponding to a distribution p majorizes the count vector $\mathbf{count}_q$ for a distribution q ($\mathbf{count}_p \succ \mathbf{count}_q$), then:*

$$AUC\text{-}C(\mathbf{count}_p) \leq AUC\text{-}C(\mathbf{count}_q)$$

Proof. The proof consists of two parts. First, we simplify the definition of AUC-C into a more direct summation over the category counts. Second, we show that this simplified form is Schur-concave using a standard theorem from the theory of majorization.

Step 1: Simplification of the AUC-C Definition The AUC-C is defined as:

$$\text{AUC-}C = \frac{1}{K_{\text{even}}} \sum_{K=0}^{K_{\text{even}}-1} C(K)$$

Let's analyze the sum $S = \sum_{K=0}^{K_{\text{even}}-1} C(K)$. We can substitute the definition of $C(K) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \mathbb{I}(\text{count}(c) > K)$, where $\mathbb{I}(\cdot)$ is the indicator function.

$$S = \sum_{K=0}^{K_{\text{even}}-1} \left(\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \mathbb{I}(\text{count}(c) > K) \right)$$

By swapping the order of summation, we obtain:

$$S = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \sum_{K=0}^{K_{\text{even}}-1} \mathbb{I}(\text{count}(c) > K)$$

The inner sum, for a specific category c, counts the number of integers K in the range $[0, K_{\text{even}} - 1]$ for which the condition $\text{count}(c) > K$ holds. Let $m_c = \text{count}(c)$. The integers K satisfying $m_c > K$ are $\{0, 1, \ldots, m_c - 1\}$. The inner sum is therefore the size of the intersection of the integer sets $\{0, 1, \ldots, m_c - 1\}$ and $\{0, 1, \ldots, K_{\text{even}} - 1\}$. The size of this intersection is $\min(m_c, K_{\text{even}})$.

Substituting this result back into the expression for S yields:

$$S = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \min(\text{count}(c), K_{\text{even}})$$

Thus, the AUC-C metric can be expressed in the simplified form:

$$\text{AUC-}C = \frac{1}{K_{\text{even}} |\mathcal{C}|} \sum_{c \in \mathcal{C}} \min(\text{count}(c), K_{\text{even}})$$

Step 2: Proof of Schur-Concavity A fundamental theorem from the theory of majorization states that a function F defined on a vector $\mathbf{x} = (x_1, \ldots, x_n)$ of the form $F(\mathbf{x}) = \sum_{i=1}^{n} g(x_i)$ is Schur-concave if and only if the function $g(x)$ is concave.

Our simplified AUC-C expression is a positive constant multiplied by a sum of this form, where the vector is the count vector $\mathbf{count} = (\text{count}(c_1), \ldots, \text{count}(c_{|\mathcal{C}|}))$, and the function g is:

$$g(x) = \min(x, K_{\text{even}})$$

To prove that AUC-C is Schur-concave, we only need to show that $g(x)$ is a concave function. A function is concave if its derivative is non-increasing. The function $g(x)$ is defined piecewise:

$$g(x) = \begin{cases} x & \text{if } x \leq K_{\text{even}} \\ K_{\text{even}} & \text{if } x > K_{\text{even}} \end{cases}$$

The derivative of $g(x)$ is:

$$g'(x) = \begin{cases} 1 & \text{if } x < K_{\text{even}} \\ 0 & \text{if } x > K_{\text{even}} \end{cases}$$

Since the derivative $g'(x)$ is non-increasing for all x, the function $g(x)$ is concave.

Because $g(x) = \min(x, K_{\text{even}})$ is concave, the summation $\sum_{c \in \mathcal{C}} g(\text{count}(c))$ is a Schur-concave function of the count vector. Multiplying by the positive constant $\frac{1}{K_{\text{even}}|\mathcal{C}|}$ preserves the Schur-concavity.

Therefore, AUC-C is a Schur-concave function. □

C Proof of Theorem 3: UCS is a Schur-concave function

Theorem 3. *The Uniform Convergence Score (UCS) is a* ***Schur-concave*** *function. This implies that if a probability distribution* $\mathbf{p}$ *majorizes a distribution* $\mathbf{q}$ *(*$\mathbf{p} \succ \mathbf{q}$*), then:*

$$UCS(\mathbf{p}) \leq UCS(\mathbf{q})$$

Proof. The Uniform Convergence Score (UCS) is defined as:

$$\text{UCS}(\mathbf{p}) = 1 - \frac{|\mathcal{C}|^2}{2(|\mathcal{C}| - 1)} \cdot A(\mathbf{p}),$$

where the term $A(\mathbf{p})$ is the area of deviation from a uniform distribution's $\overline{C}(Q)$ curve:

$$A(\mathbf{p}) = \int_0^{q^*} (1 - \overline{C}(Q; \mathbf{p}))\, dQ + \int_{q^*}^1 \overline{C}(Q; \mathbf{p})\, dQ.$$

Let $K = \frac{|\mathcal{C}|^2}{2(|\mathcal{C}|-1)}$ be the positive normalization constant. The definition is thus $\text{UCS}(\mathbf{p}) = 1 - K \cdot A(\mathbf{p})$.

To prove that $\text{UCS}(\mathbf{p})$ is a Schur-concave function, we must show that for any two probability vectors $\mathbf{p}$ and $\mathbf{q}$ of the same dimension, if $\mathbf{p} \succ \mathbf{q}$ (p majorizes q), then $\text{UCS}(\mathbf{p}) \leq \text{UCS}(\mathbf{q})$.

From the definition, the inequality $\text{UCS}(\mathbf{p}) \leq \text{UCS}(\mathbf{q})$ is equivalent to:

$$1 - K \cdot A(\mathbf{p}) \leq 1 - K \cdot A(\mathbf{q})$$

$$-K \cdot A(\mathbf{p}) \leq -K \cdot A(\mathbf{q})$$

Since $K > 0$, this simplifies to:

$$A(\mathbf{p}) \geq A(\mathbf{q})$$

This means that proving $\text{UCS}(\mathbf{p})$ is Schur-concave is equivalent to proving that the area of deviation, $A(\mathbf{p})$, is a **Schur-convex** function.

A function is Schur-convex if its value increases or stays the same when its input vector becomes less uniform (in the sense of majorization). The term $A(\mathbf{p})$ is a measure of statistical dispersion or inequality; it quantifies the total deviation of the $\overline{C}(Q; \mathbf{p})$ curve from the reference curve of a perfectly uniform distribution.

Such measures of dispersion are archetypal examples of Schur-convex functions. The intuition is as follows: the majorization order $\mathbf{p} \succ \mathbf{q}$ is fundamentally defined by transformations (known as T-transforms) that take two distinct components of $\mathbf{q}$, say $q_i > q_j$, and move them further apart to create $p_i = q_i + \delta$ and $p_j = q_j - \delta$ for some $\delta > 0$. This operation makes the distribution less uniform.

Applying such a transform to the probability vector has a direct effect on the $\overline{C}(Q)$ curve. The locations of the steps in the curve are the probability values themselves. Moving two of these values, q_i and q_j, further apart effectively "stretches" the $\overline{C}(Q)$ curve horizontally. This stretching increases the integrated absolute difference between the curve and the fixed reference curve of the uniform distribution. Therefore, for any such transform that makes a distribution less uniform, the area $A(\mathbf{p})$ will increase or stay the same.

Thus, $A(\mathbf{p})$ is a Schur-convex function. It follows that $-K \cdot A(\mathbf{p})$ is Schur-concave. Adding a constant, 1, does not alter the concavity property.

We conclude that UCS($\mathbf{p}$) is a Schur-concave function. □

D Proof of Proposition 4: Linear Relationship between UCS and TVD to Uniform

Proposition 5 (Linear Relationship between UCS and TVD to Uniform)**.** *Let* $\mathbf{p} = (p_1, \ldots, p_{|\mathcal{C}|})$ *be a probability distribution over a set of* $|\mathcal{C}|$ *categories, and let* $\mathbf{u}$ *be the uniform distribution over the same categories, where* $u_i = 1/|\mathcal{C}|$ *for all* i*.*

The Uniform Convergence Score (UCS) of $\mathbf{p}$ *is a direct linear transformation of the Total Variation Distance (TVD) between* $\mathbf{p}$ *and* $\mathbf{u}$*. Their relationship is given by the exact formula:*

$$UCS(\mathbf{p}) = 1 - \frac{|\mathcal{C}|}{|\mathcal{C}| - 1} \cdot TVD(\mathbf{p}, \mathbf{u})$$

Proof. We begin with the definition of UCS, which can be expressed as a function of the L1 distance to the uniform distribution:

$$\text{UCS}(\mathbf{p}) = 1 - \frac{|\mathcal{C}|}{2(|\mathcal{C}| - 1)} \sum_{i=1}^{|\mathcal{C}|} |p_i - 1/|\mathcal{C}||$$

The definition of the Total Variation Distance between $\mathbf{p}$ and the uniform distribution $\mathbf{u}$ is:

$$\text{TVD}(\mathbf{p}, \mathbf{u}) = \frac{1}{2} \sum_{i=1}^{|\mathcal{C}|} |p_i - u_i| = \frac{1}{2} \sum_{i=1}^{|\mathcal{C}|} |p_i - 1/|\mathcal{C}||$$

From the definition of TVD, we can express the summation term (the L1 distance) as:

$$\sum_{i=1}^{|\mathcal{C}|} |p_i - 1/|\mathcal{C}|| = 2 \cdot \text{TVD}(\mathbf{p}, \mathbf{u})$$

Substituting this into our expression for UCS:

$$\text{UCS}(\mathbf{p}) = 1 - \frac{|\mathcal{C}|}{2(|\mathcal{C}| - 1)} (2 \cdot \text{TVD}(\mathbf{p}, \mathbf{u}))$$

Canceling the factor of 2, we arrive at the final relationship:

$$\text{UCS}(\mathbf{p}) = 1 - \frac{|\mathcal{C}|}{|\mathcal{C}| - 1} \cdot \text{TVD}(\mathbf{p}, \mathbf{u})$$

This demonstrates that UCS is a linear transformation of TVD. As a consequence, they will produce the exact same ranking of distributions in terms of uniformity. □