

**A Poisson Count Race as a Generative Bridge Between Logit and Probit Choice**

Vencislav Popov

Department of Psychology, University of Zurich

**Author Note**

Vencislav Popov  <https://orcid.org/0000-0002-8073-4199>

Correspondence concerning this article should be addressed to Vencislav Popov,  
Email: [vencislav.popov@gmail.com](mailto:vencislav.popov@gmail.com)

### Abstract

This study unifies the two canonical discrete choice specifications - multinomial logit and multinomial probit - within a single count process. The model uses a Poisson count process race, wherein  $K$  alternatives generate events according to independent Poisson processes. A choice is determined when an alternative reaches a cumulative count threshold  $\theta$ . At the single-event threshold,  $\theta = 1$ , the model yields the Multinomial Logit (Luce choice rule). By normalizing the utility noise to maintain a constant variance, I show that as  $\theta \rightarrow \infty$ , the model converges to the Multinomial Probit. This formulation provides a parametric bridge in which  $\theta$  governs the shape of the error distribution and a separate parameter,  $\beta$ , modulates the variance, while the systematic utilities  $v_i$  are shared across all regimes. The generalized model uses a single stochastic accumulation mechanism: both regimes belong to the log-Gamma random utility family, with  $\theta$  governing the transition between Gumbel and Gaussian noise shapes. The unification is distributional rather than dynamical: changes in  $\theta$  do not represent a single decision-making system adjusting its threshold in real time, but rather compare different accumulation regimes at matched discriminability.

*Keywords:* discrete choice, random utility, multinomial logit, multinomial probit, Poisson process, evidence accumulation

### A Poisson Count Race as a Generative Bridge Between Logit and Probit Choice

Discrete choice models based on Random Utility (RU) theory (McFadden, 1974) form a central pillar of mathematical psychology, econometrics, and cognitive science. In these models, each alternative in a choice set elicits a *latent* scalar quantity - often interpreted as strength, utility, or evidence - and the observed choices arise from a comparison of these latent quantities under stochastic variability. The general setup is surprisingly simple. Consider a decision-maker facing a set of  $K \geq 2$  mutually exclusive alternatives. The utility associated with alternative  $i$  can be decomposed into a systematic component,  $v_i$ , and a stochastic component,  $\epsilon_i$ :

$$U_i = v_i + \epsilon_i, \quad i = 1, \dots, K$$

The decision-maker then selects the alternative with maximum utility:

$$C = \arg \max_{i \in K} U_i$$

While the term “utility” implies an economic setting, the underlying mathematical model concerns any situation in which choices are probabilistic. The core concept is that of a shared *random scale* on which multiple latent variables, each associated with a different choice, are represented. Any choice model that implements this assumption, together with the max choice rule, is said to have a random-scale representation or to be random-scale representable (Falmagne, 1978; Kellen et al., 2021).

Within this general framework, different assumptions about the distribution of errors,  $\epsilon_i$ , determine the structure of the choice model and its predictions. Two important classes dominate the field: logit-based models, derived from Luce’s choice axiom and extreme-value theory (Luce, 1959; McFadden, 1974; Yellott, 1977), and probit-based models, derived from Thurstone’s Theory of comparative judgements and gaussian Signal Detection Theory (Hausman & Wise, 1978; Robinson et al., 2023; Thurstone, 1927; Wixted, 2020). These two model families correspond to two types of error distributions:

1. **Gumbel Errors:** If the  $\epsilon_i$  terms are independent and identically distributed (i.i.d.)

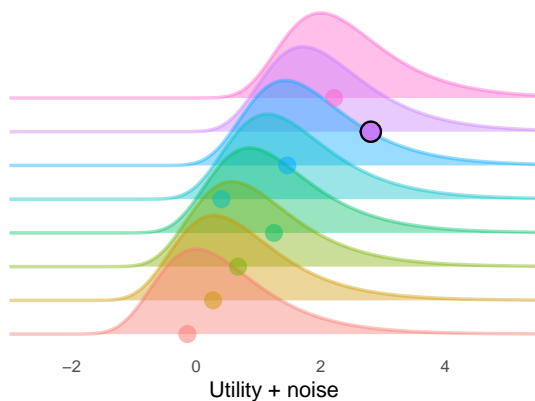
according to a Type I Extreme Value distribution, the choice probabilities follow the **Multinomial Logit (MNL)** or Softmax form.

2. **Gaussian Errors:** If the  $\epsilon_i$  terms follow a Multivariate Normal distribution (potentially allowing for correlated errors across alternatives) the choice probabilities are described by the **Multinomial Probit (MNP)** model.

**Figure 1**

The two canonical random utility models. Left panels show the noise distributions for each alternative, with a single random draw from each (the winning draw is highlighted). Right panels show the corresponding choice probability formulas. Top row: Gumbel (Type I Extreme Value) noise yields the Multinomial Logit (Softmax). Bottom row: Gaussian noise yields the Multinomial Probit.

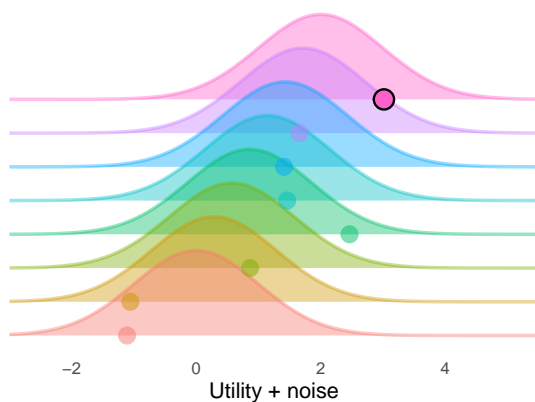
Gumbel (Type I EV) noise



Multinomial Logit (Softmax)

$$P(C = i) = \frac{\exp(v_i)}{\sum_{j=0}^K \exp(v_j)}$$

Gaussian (Normal) noise



(Independent) Multinomial Probit

$$P(C = i) = \int_{-\infty}^{+\infty} \prod_{j \neq i} \Phi(x - v_j) \cdot \phi(x - v_i) dx$$

The random utility framework thus unifies axiomatic choice models (Luce, 1959) and measurement detection-based models (Thurstone, 1927) under a single functional form.

However, the unification currently stops there - the different error distributions reflect different assumptions about the generating process. As Robinson et al. (2023) recently put it, the two models “describe different ways of translating sensory evidence into decision variables.”.

The central question addressed in this paper is this: Can the two canonical random-utility discrete-choice specifications — multinomial logit and multinomial probit — be derived from a shared generative mechanism? Can we find a deeper unifying principle? The surprising answer is yes - both models can be derived as limit cases of a single stochastic evidence accumulation mechanism. The key distinction between the models is not the form of the utility noise itself, but the stopping rule governing how stochastic evidence is accumulated before a choice is made.

### 0.1 Model overview

Take a **Poisson count race** (Pike, 1973; Smith & Van Zandt, 2000; Townsend & Ashby, 1983), wherein each alternative generates stochastic events via an independent Poisson process. A decision is made when one alternative reaches a cumulative count threshold  $\theta$ . The threshold  $\theta$  thus controls how much stochastic evidence must be accumulated before commitment.

Without any additional modifications, such a model predicts increasingly deterministic responses, and in the limit as  $\theta \rightarrow \infty$  it leads to choosing the option with the highest utility 100% of the time. To prevent this, and to compare noise *shape* across  $\theta$  independently of noise *scale*, we must standardize the stochastic utility component to unit variance for all  $\theta$ , with a separate parameter  $\beta$  controlling the overall magnitude of stochasticity. With this normalization, we can establish three main results:

1. At  $\theta = 1$ , this reduces exactly to the Multinomial Logit
2. For any  $\theta \geq 1$ , the Poisson count race is isomorphic to a random utility model with log-Gamma noise, interpolating between Gumbel and Gaussian error distributions;
3. As  $\theta \rightarrow \infty$  the model converges to the Multinomial Probit.

From this perspective, logit and probit can be understood as members of a single parametric family of evidence accumulation models that differ in the accumulation stopping

rule. Extreme-value noise and Gaussian noise arise as the two endpoint regimes of log-Gamma noise, which is the natural error distribution of Poisson count races. It is important to note, however, that this unification is algebraic and distributional rather than dynamical: the standardization that enables comparison across  $\theta$  values compares different accumulation systems at matched discriminability, not a single system under threshold variation (see Section 3 for details).

While Poisson counter models have a rich history in mathematical psychology—particularly as accounts of response time distributions under time-varying evidence rates (Pike, 1973; Smith & Van Zandt, 2000; Townsend & Ashby, 1983) - they have rarely been invoked to address the theoretical relationship between *static* discrete choice models. For example, Smith and Van Zandt (2000) showed that Poisson races yield choice probabilities governed by the incomplete Beta function when there are two options to choose from; but their analysis focused on relatively small integer thresholds  $\theta \approx 5\text{--}10$  chosen to capture reaction-time skewness.

The present work takes a different perspective. Rather than treating the threshold as a fixed descriptive parameter, I examine the asymptotic behavior of the Poisson count race as  $\theta \rightarrow \infty$  under variance-preserving identification. This shift in emphasis reveals that the Poisson race is not merely a model of latency, but a generative mechanism that continuously interpolates between the two canonical pillars of discrete choice: Multinomial Logit  $\theta = 1$  and Multinomial Probit  $\theta \rightarrow \infty$ .

The remainder of the paper formalizes this framework, presents simulations illustrating the interpolation between regimes, and discusses implications for discrete choice modeling.

## 1 The Generative Model: A Poisson Count Race

Let the accumulation of evidence or preference for each alternative  $i$  be modeled by independent Poisson count processes, denoted  $N_i(t)$ , with rate parameters  $\lambda_i > 0$ . Here  $N_i(t)$  represents cumulative count for alternative  $i$  at time  $t$ .

Then, define a **count race** characterized by an integer threshold  $\theta \geq 1$ . The process terminates as soon as any single alternative accumulates  $\theta$  events.

**Definition 1 (Stopping Time).** The stopping time for the system is the first time any

process hits the threshold:

$$\tau_\theta = \inf\{t \geq 0 : \max_i N_i(t) = \theta\}$$

**Definition 2 (Choice).** The chosen alternative is the specific process that triggers the stopping time:

$$C = \arg \max_i N_i(\tau_\theta)$$

### 1.1 Transformation to Waiting Times

To map this stochastic process to a random utility framework, consider  $T_i^{(\theta)}$ , the waiting time until the  $i$ -th process records its  $\theta$ -th event:

$$T_i^{(\theta)} := \inf\{t : N_i(t) = \theta\}$$

For a Poisson process with rate  $\lambda_i$ , the waiting time to the  $\theta$ -th jump follows a Gamma (Erlang) distribution with shape  $\theta$  and rate  $\lambda_i$ :

$$T_i^{(\theta)} \sim \text{Gamma}(\text{shape} = \theta, \text{rate} = \lambda_i)$$

The condition that alternative  $i$  wins the race is equivalent to observing the minimum waiting time:

$$C = \arg \min_i T_i^{(\theta)}$$

### 1.2 The Random Utility Representation

Since the Poisson processes are independent, the waiting times  $T_1^{(\theta)}, \dots, T_K^{(\theta)}$  are mutually independent. Utilizing the scaling property of the Gamma distribution, we can express each waiting time as:

$$T_i^{(\theta)} \stackrel{d}{=} \frac{G_i}{\lambda_i}$$

where  $G_1, \dots, G_K \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\theta, 1)$  are standard Gamma random variables. The choice problem then becomes:

$$C = \arg \min_i \left( \frac{G_i}{\lambda_i} \right)$$

Applying the natural logarithm and taking the negative, a monotonic transformation, reverses the optimization direction from minimization to maximization:

$$\begin{aligned} C &= \arg \min_i (\log G_i - \log \lambda_i) \\ &= \arg \max_i (\log \lambda_i - \log G_i) \end{aligned}$$

This establishes an exact Random Utility Model (RUM) structure:

$$U_i^{(\theta)} = v_i + \epsilon_i^{(\theta)}$$

where:

- **Systematic Utility:**  $v_i = \log \lambda_i$
- **Stochastic Error:**  $\epsilon_i^{(\theta)} = -\log G_i$ , with  $G_i \sim \text{Gamma}(\theta, 1)$ .

Thus, the Poisson count race is isomorphic to a Random Utility Model characterized by Log-Gamma noise.

#### Result 1 (Log-Gamma Random Utility Family)

For any integer threshold  $\theta \geq 1$ , the Poisson count race induces a random utility model  $U_i^{(\theta)} = \log \lambda_i - \log G_i$  with i.i.d. log-Gamma( $\theta$ ) noise.

## 2 The Logit Boundary ( $\theta = 1$ )

In the specific instance where the threshold is a single event ( $\theta = 1$ ), the waiting time distribution simplifies to the Exponential distribution:

$$G_i \sim \text{Gamma}(1, 1) \equiv \text{Exponential}(1)$$

A fundamental property of Extreme Value Theory is the relationship between the Exponential and Gumbel distributions:

$$X \sim \text{Exp}(1) \implies -\log(X) \sim \text{Gumbel}(\text{Type I EV})$$

Therefore, when  $\theta = 1$ , the noise terms  $\epsilon_i^{(1)}$  are i.i.d. standard Gumbel. This recovers the exact Multinomial Logit formula:

$$\Pr(C = i) = \frac{\exp(v_i)}{\sum_{j=1}^K \exp(v_j)} = \frac{\lambda_i}{\sum_{j=1}^K \lambda_j}$$

#### Result 2 (Logit Boundary)

The Poisson count race with threshold  $\theta = 1$  recovers the Multinomial Logit model exactly, with deterministic utility components equal to the log rate of their poisson counterparts. The accumulation of a single unit of evidence corresponds to the Luce Choice Rule (Soft-max). This is a classic, well-known derivation.

### 3 Variance normalization

For thresholds  $\theta > 1$ , the error distribution deviates from the Gumbel form. More critically, as  $\theta$  increases, the variance of the error term diminishes. Specifically, for  $\epsilon^{(\theta)} = -\log G$  where  $G \sim \text{Gamma}(\theta, 1)$ , the moments are:

$$\mathbb{E}[\epsilon^{(\theta)}] = -\psi(\theta)$$

$$\text{Var}(\epsilon^{(\theta)}) = \psi_1(\theta)$$

where  $\psi(\cdot)$  is the digamma function and  $\psi_1(\cdot)$  is the trigamma function.

As  $\theta \rightarrow \infty$ , the variance  $\psi_1(\theta) \approx 1/\theta \rightarrow 0$ . Without intervention, the model would converge to a deterministic choice rule (argmax of systematic utilities) simply because the noise vanishes. To facilitate a meaningful comparison of error shapes across varying  $\theta$ , we must enforce a consistent scale.

Define the standardized noise term  $Z_i^{(\theta)}$  to have zero mean and unit variance for all  $\theta$ :

$$Z_i^{(\theta)} := \frac{\epsilon_i^{(\theta)} - \mu_\theta}{\sigma_\theta} = \frac{-\log G_i + \psi(\theta)}{\sqrt{\psi_1(\theta)}}$$

This leads to family of utility models with matched discriminability:

$$U_i^{(\theta)} = v_i + \beta Z_i^{(\theta)}$$

Here:

- $v_i$  is the systematic utility (evidence rate).
- $\theta$  governs the shape of the noise (from skewed Gumbel to symmetric Gaussian).
- $\beta$  governs the temperature (the magnitude of noise relative to utility).

An important caveat accompanies this construction. In a random utility model, rescaling all utilities by a common positive constant does not change choice probabilities. If we multiply both terms by  $\sigma_\theta = \sqrt{\psi_1(\theta)}$ , we would not change the predicted probabilities. Therefore, the variance-normalized model  $U_i^{(\theta)} = v_i + \beta Z_i^{(\theta)}$  is observationally equivalent to a model with systematic utilities  $v_i \sigma_\theta$  and *unstandardized* log-Gamma noise  $\epsilon_i^{(\theta)}$ .

This has an important theoretical consequence: the standardized distributions cannot describe the behavior of a single accumulation system under threshold variation. For a fixed set of Poisson rates  $\lambda_i$ , increasing  $\theta$  both reshapes the noise and reduces its variance; the variance reduction alone drives choice toward determinism regardless of shape. The standardization removes this confound by comparing across systems with different effective rate structures — specifically, the effective rates would need to scale as  $\lambda_i^{\sigma_\theta}$  to maintain constant discriminability. However, it is not psychologically plausible that a decision-maker can affect the utilities / poisson rates in exactly the right way if they choose a higher, more conservative threshold. Simply put - a single decision-making system cannot wait its way from a logit to a probit model while keeping discriminability fixed.

The unification established here is therefore *distributional* — logit and probit belong to the same parametric family of log-Gamma random utility models — rather than *dynamical*. One cannot convert a logit-like decision process into a probit-like one merely by raising the decision threshold within a single system. Rather, different regimes likely describe the functioning of different decision-making systems.

#### 4 The Probit Limit ( $\theta \rightarrow \infty$ )

This section establishes the asymptotic behavior of the variance-standardized model.

Recall that  $G_i \sim \text{Gamma}(\theta, 1)$ . For integer  $\theta$ ,  $G_i$  can be represented as the sum of  $\theta$  independent exponential variables:  $G_i = \sum_{j=1}^{\theta} E_{ij}$ . By the Central Limit Theorem, the standardized variable converges to a standard normal distribution:

$$\frac{G_i - \theta}{\sqrt{\theta}} \xrightarrow{d} \mathcal{N}(0, 1)$$

We are interested in the distribution of the log-transformed variable,  $\epsilon_i^{(\theta)} = -\log G_i$ . By applying the Delta Method with the transformation  $g(x) = -\log x$ , the asymptotic distribution of  $-\log G_i$  is normal with variance  $[g'(\theta)]^2 \cdot \text{Var}(G_i) = (-1/\theta)^2 \cdot \theta = 1/\theta$ .

Standardizing this result matches our variance standardization scaling. Since  $\psi_1(\theta) \sim 1/\theta$  for large  $\theta$ , the standardized term converges to the standard normal:

$$Z_i^{(\theta)} \xrightarrow{d} \mathcal{N}(0, 1)$$

Consequently, the vector of random utilities converges in distribution:

$$(v_i + \beta Z_i^{(\theta)})_{i=1}^K \xrightarrow{d} (v_i + \beta Z_i)_{i=1}^K$$

where  $Z_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ .

### Result 3 (Probit Limit)

As  $\theta \rightarrow \infty$ , the Variance-Standardized Poisson race converges to the Independent Multinomial Probit model. Within the variance-standardized family, Logit corresponds to the  $\theta = 1$  member and Probit to the  $\theta \rightarrow \infty$  limit. This characterizes the two models as occupying different positions within a single parametric family of log-Gamma random utility models, indexed by the accumulation threshold.

## 5 Simulation Studies: Binary Choice

To illustrate how the Poisson count race family interpolates between Logit and Probit, I conducted a simulation study in the binary choice setting ( $K = 2$ ). This setting admits closed-form choice probabilities and allows direct visual comparison with both classical models.

Let  $\lambda_1$  and  $\lambda_2$  denote the Poisson rates of the two alternatives, and define the log-rate ratio  $x = \log(\lambda_1/\lambda_2)$ . The probability that alternative 1 wins the race can be expressed in closed form using the regularized incomplete Beta function:

$$\Pr(C = 1 \mid x, \theta) = I_{\sigma(x)}(\theta, \theta)$$

where  $I_p(a, b)$  is the regularized incomplete Beta function and  $\sigma(x) = (1 + e^{-x})^{-1}$ . To see this, note that alternative 1 wins the race if and only if  $T_1^{(\theta)} < T_2^{(\theta)}$ , where  $T_i^{(\theta)} \sim \text{Gamma}(\theta, \lambda_i)$ . Equivalently, defining  $W = T_1^{(\theta)} / (T_1^{(\theta)} + T_2^{(\theta)})$ , we have  $W \sim \text{Beta}(\theta, \theta)$  when evaluated at  $p = \lambda_1 / (\lambda_1 + \lambda_2) = \sigma(x)$ , giving  $\Pr(T_1 < T_2) = I_{\sigma(x)}(\theta, \theta)$ . For  $\theta = 1$ , this reduces exactly to  $\sigma(x)$ , the logistic function.

When choice probabilities are plotted directly as a function of  $x$  for increasing  $\theta$ , the choice function becomes increasingly steep and converges to a step function at  $x = 0$ , reflecting deterministic selection of the alternative with the larger rate. This confirms that without variance standardization, increasing the count threshold simply reduces stochasticity rather than inducing Gaussian behavior.

To compare noise *shape* independently of noise *scale*, I adopted the variance standardization introduced in Section 3. For binary choice, the variance of the utility noise difference is  $\text{Var}(\epsilon_1^{(\theta)} - \epsilon_2^{(\theta)}) = 2\psi_1(\theta)$ . I therefore define the standardized signal  $s = x / \sqrt{2\psi_1(\theta)}$ . On this variance-matched axis, the Logit reference ( $\theta = 1$ ) uses  $\text{sd}_{\text{diff}} = \pi / \sqrt{3}$  (the standard deviation of the difference of two independent Gumbel variates), and the Probit reference is simply  $\Phi(s)$ .

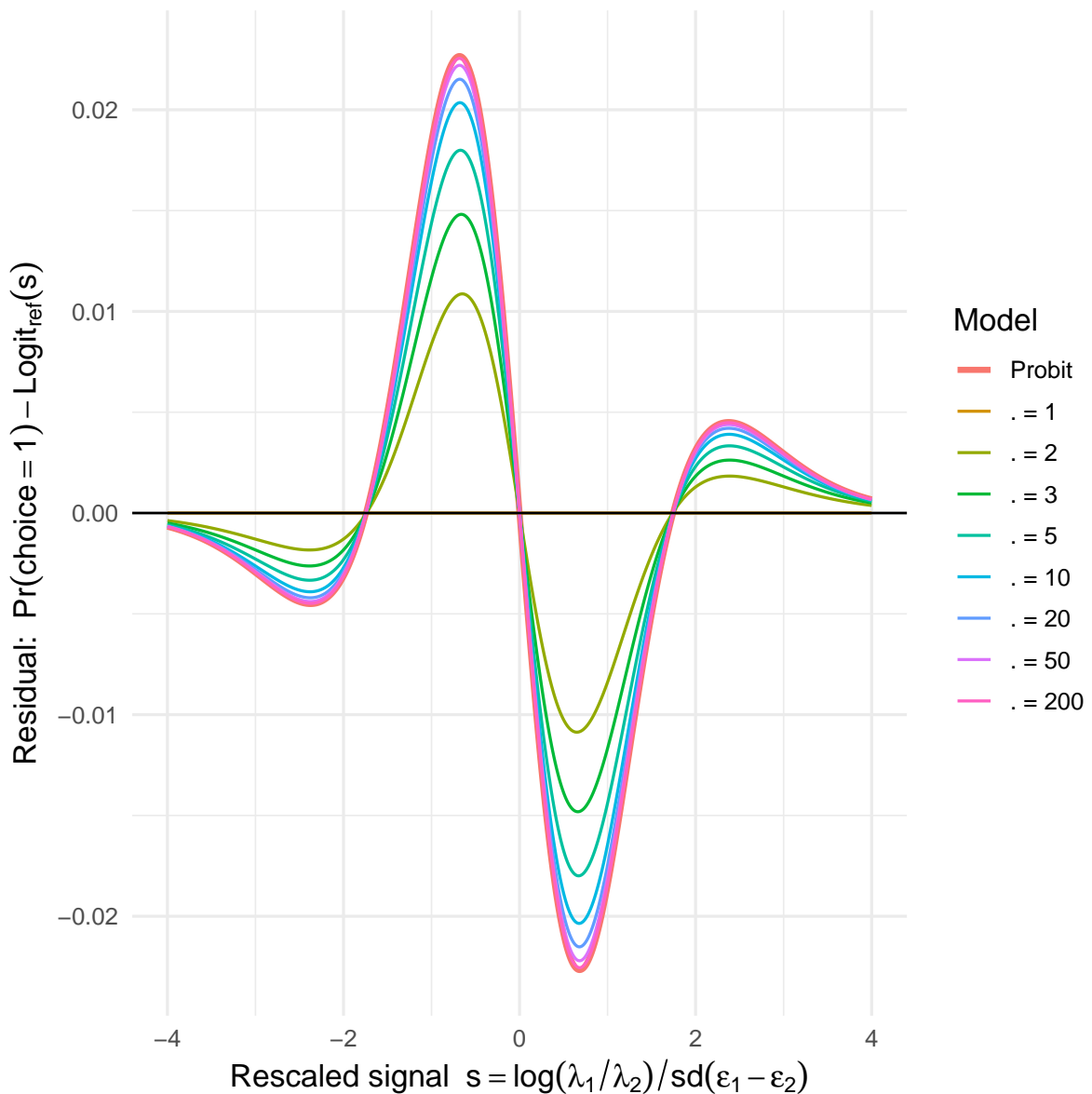
For the multinomial simulations, the same principle applies. The logit reference uses an effective inverse temperature of  $\pi / (\beta\sqrt{6})$ , which ensures the Gumbel noise has standard deviation matching  $\beta$  under the unit-variance convention. The probit reference uses noise scale  $\beta$  directly. All simulations use  $\beta = 1$  unless otherwise noted. Under this normalization, the  $\theta = 1$  Poisson race coincides exactly with the variance-matched logit curve, while increasing  $\theta$  yields choice functions that converge uniformly to the probit curve.

Because logit and probit are themselves numerically close under variance matching, the differences between models are small in absolute magnitude but systematic. To make these differences visible, Figure 2 plots residuals relative to the variance-matched Logit model. At  $\theta = 1$ , the residual is identically zero (exact Logit). As  $\theta$  increases, the residuals grow smoothly and converge toward the Probit–Logit difference curve, with the maximum absolute

deviation from probit decaying rapidly in  $\theta$ . This confirms that the variance-standardized Poisson count race defines a continuous, parameterized family of choice rules that interpolates smoothly between logit-like and probit-like behavior.

**Figure 2**

*Residuals of Poisson count race choice probabilities relative to the Logit reference, plotted on a variance-matched axis. Each curve corresponds to a different accumulation threshold  $\theta$ . At  $\theta = 1$  the model is exactly Logit (zero residual). As  $\theta$  increases, the curves converge toward the Probit – Logit difference (black curve), confirming the theoretical bridge between the two models.*



## 6 Simulation Studies: Multinomial $K$ -alternative Choice

The binary simulation demonstrates that the variance-standardized Poisson count race interpolates smoothly between Logit ( $\theta = 1$ ) and Probit ( $\theta \rightarrow \infty$ ) in the two-alternative case. Here I extend this analysis to the multinomial setting ( $K > 2$ ), where the differences between logit and probit become richer and more consequential.

In the binary case, logit and probit choice functions differ only in the shape of the psychometric curve—a subtle quantitative distinction. With three or more alternatives, additional qualitative differences emerge. Most prominently, the Multinomial Logit model satisfies the *Independence of Irrelevant Alternatives* (IIA) property: the ratio of choice probabilities for any two alternatives is independent of the remaining alternatives in the choice set (Luce, 1959). The Multinomial Probit model, even with independent errors, does not share this property (Hausman & Wise, 1978). The Poisson count race therefore provides a window into how IIA-like behavior gradually weakens as the noise distribution transitions from Gumbel to Gaussian.

The multinomial simulations are organised around five questions:

1. **Convergence:** How quickly do Poisson count race choice probabilities converge to the Probit reference as  $\theta$  increases, and does the rate of convergence depend on  $K$ ?
2. **Probability vectors:** How does the full distribution over alternatives change as  $\theta$  varies from 1 to large values?
3. **Set-size scaling:** How does the probability of choosing a target alternative scale with the number of competitors, and how does this scaling differ between logit, probit, and intermediate regimes?
4. **Independence of Irrelevant Alternatives:** How does the IIA property—exact under logit—erode as  $\theta$  increases toward the probit regime?
5. **Parameter invariance:** When misspecified logit or probit models are fit to Poisson count race data, which model yields parameters that are invariant to  $K$ ?

All simulations use Monte Carlo sampling with  $10^6$  to  $10^7$  replications per condition unless otherwise noted.

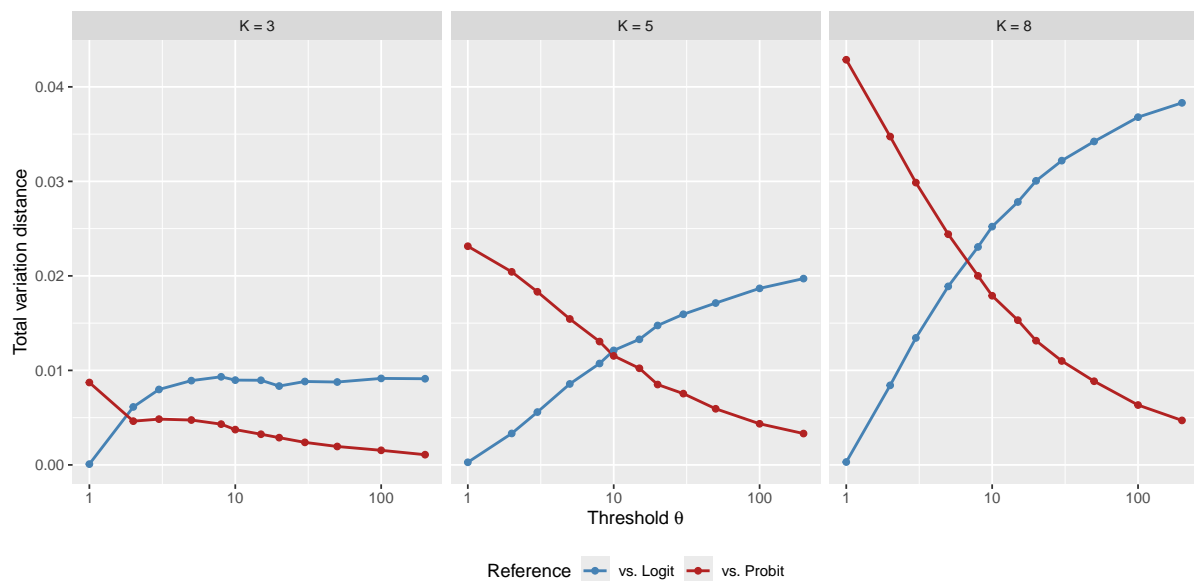
### 6.1 Study 1: Convergence to Probit

I first examine how the total variation (TV) distance between the Poisson count race choice probabilities and the Logit / Probit references changes as a function of  $\theta$ , for different numbers of alternatives  $K$ .

For each  $K$ , I use linearly spaced utilities  $v_i = (K - i)/(K - 1)$  for  $i = 1, \dots, K$ , ensuring that the best and worst alternatives always have utilities 1 and 0 regardless of  $K$ . The temperature is fixed at  $\beta = 1$ .

**Figure 3**

*Total variation distance between Poisson count race choice probabilities and the Logit (blue) and Probit (red) references, as a function of threshold  $\theta$ , for different numbers of alternatives  $K$ .*

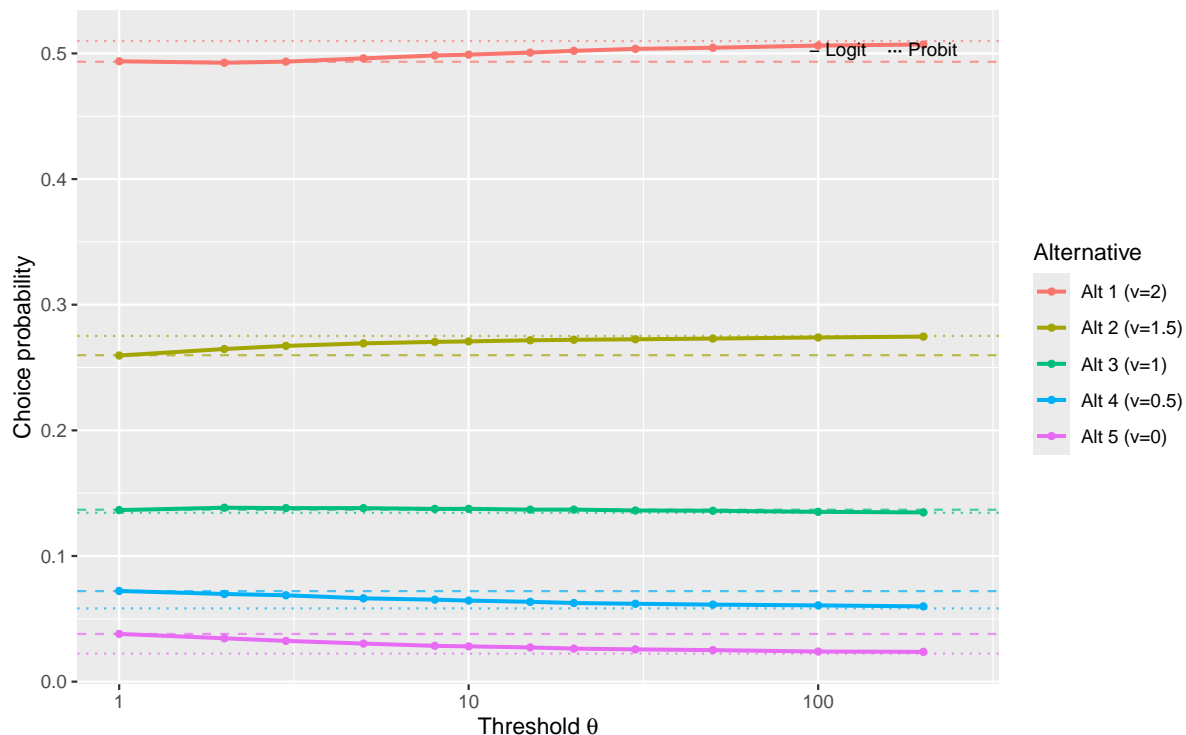


### 6.2 Study 2: Choice Probability Vectors

To visualise how the full distribution over alternatives evolves with  $\theta$ , I fix  $K = 5$  with utilities  $v = (2.0, 1.5, 1.0, 0.5, 0.0)$  and plot the choice probability for each alternative across a range of thresholds.

**Figure 4**

Choice probabilities for each of five alternatives as a function of threshold  $\theta$ . Dashed lines: Logit reference; dotted lines: Probit reference.



### 6.3 Study 3: Set-Size Scaling

A critical diagnostic for discriminating between logit and probit models is the effect of adding alternatives to the choice set (Robinson et al., 2023). Under MNL, the probability of choosing a target alternative with fixed utility is strictly determined by the ratio of its strength to the total strength. Under MNP, the scaling with set size differs because the probability of “winning” the maximum comparison depends on the shape of the noise distribution.

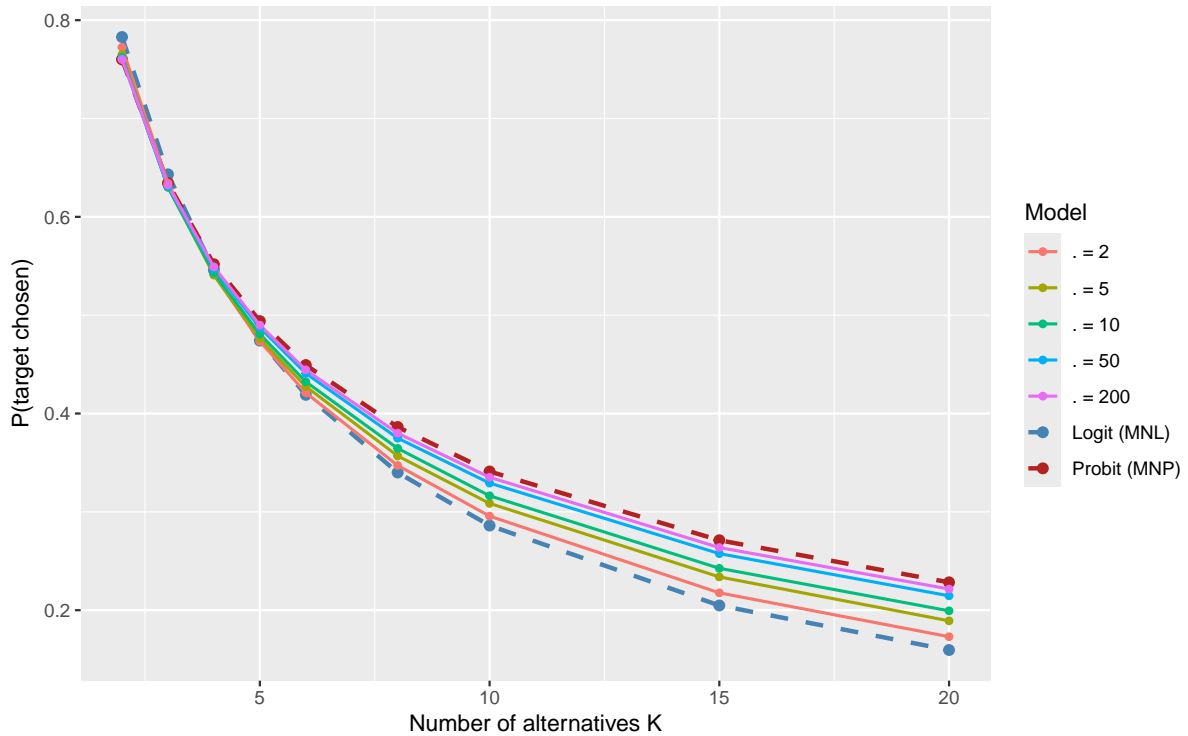
I fix a target alternative with utility  $v_{\text{target}} = 1$  and add  $K - 1$  equal competitors, each with utility  $v_{\text{comp}} = 0$ . As  $K$  grows, I track the probability of choosing the target.

Under MNL, the choice probability is  $P(\text{target}) = e^a / (e^a + K - 1)$ , where  $a = v_t \cdot \pi / (\beta\sqrt{6})$  is the effective scaled utility.

Under MNP:  $P(\text{target}) = \int \phi(z) \Phi(v_t/\beta + z)^{K-1} dz$  (by symmetry of the  $K - 1$  equal competitors). This integral reveals that probit’s thinner tails give the target a larger advantage over many competitors than logit’s heavier tails.

**Figure 5**

Probability of choosing a target alternative ( $v = 1$ ) against  $K - 1$  equal competitors ( $v = 0$ ) as a function of set size  $K$ . MNL (dashed blue) and MNP (dashed red) references are shown.



#### 6.4 Study 4: Independence of Irrelevant Alternatives

The IIA property is a hallmark of the Multinomial Logit model: the ratio of choice probabilities for any two alternatives is invariant to the composition of the choice set.

Formally, for alternatives  $i$  and  $j$ :

$$\frac{P(i | C)}{P(j | C)} = \frac{e^{v_i}}{e^{v_j}} \quad \text{for all choice sets } C \ni i, j$$

This property does not hold for the Multinomial Probit model, even when errors are independent and identically distributed. The Poisson count race therefore provides a mechanism through which IIA holds exactly at  $\theta = 1$  and is progressively violated as  $\theta$  increases.

To quantify this, I consider three alternatives with utilities  $v = (2, 1, 0)$ . I compute the ratio  $P(1)/P(2)$  under two conditions:

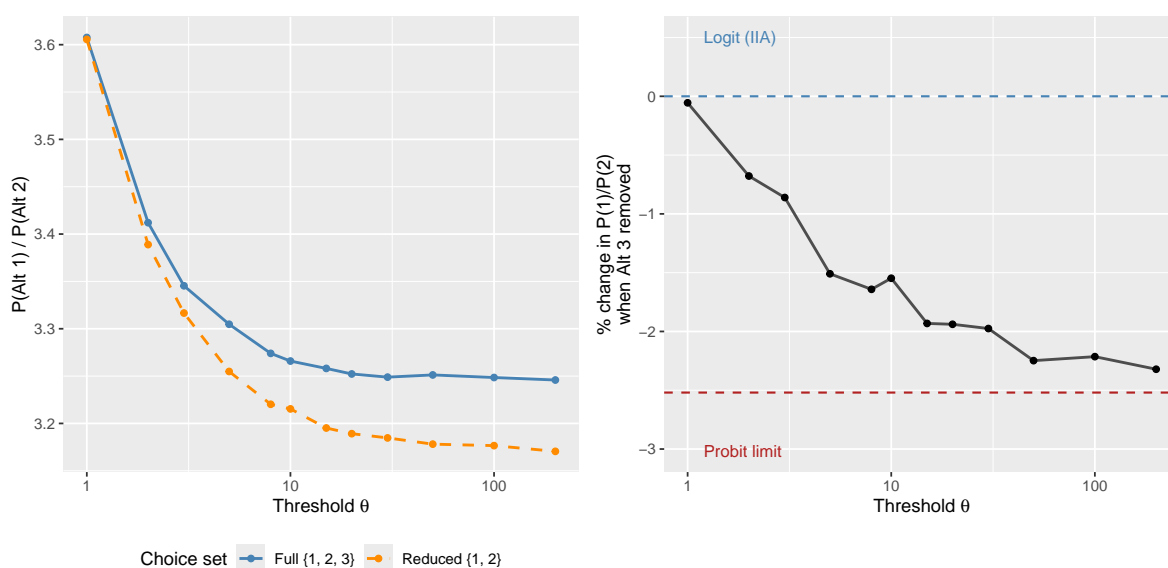
- **Full set:** all three alternatives present  $\{1, 2, 3\}$

- **Reduced set:** only alternatives {1, 2} present

Under IIA, these ratios should be identical. I track the percentage change in the ratio as  $\theta$  varies.

**Figure 6**

*Test of IIA. Left: the ratio  $P(\text{Alt } 1)/P(\text{Alt } 2)$  in the full set {1, 2, 3} (solid) and reduced set {1, 2} (dashed). Right: percentage change in the ratio when alternative 3 is removed.*



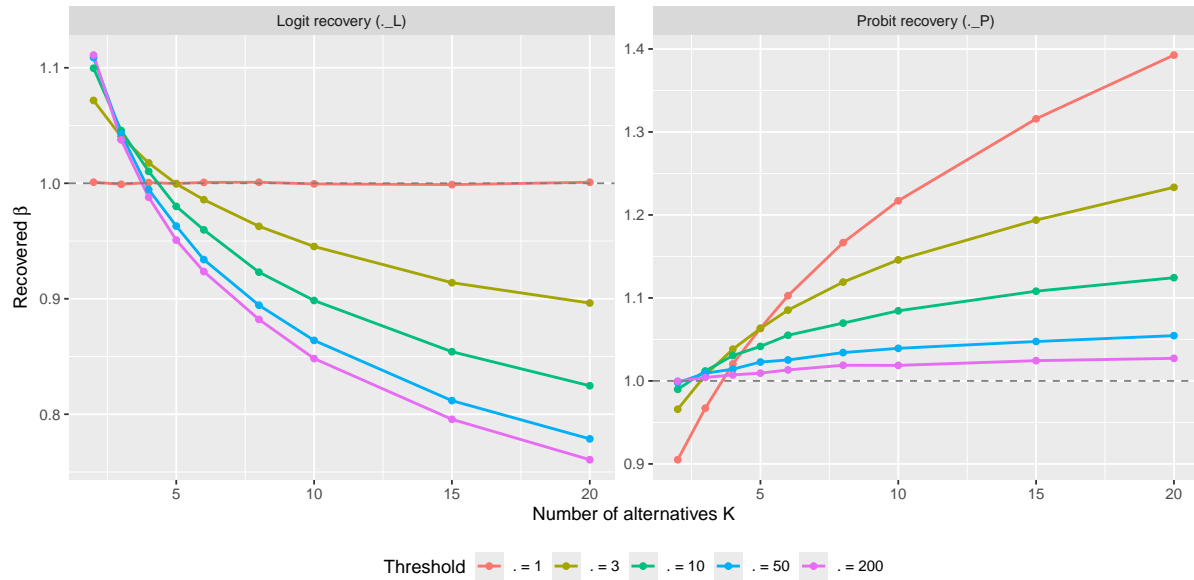
### 6.5 Study 5: Parameter Invariance Across Set Size

A key empirical diagnostic for distinguishing between logit and probit is parameter invariance across changes in set size  $K$  (Robinson et al., 2023). If choice data are generated by a logit model, the softmax inverse temperature  $\beta_{\text{logit}}$  recovered from fitting a logit specification should remain constant as  $K$  increases. Conversely, if the data follow a probit model, the Gaussian noise scale  $\beta_{\text{probit}}$  should be invariant to  $K$ .

I test this directly. For each value of  $\theta$  and each set size  $K$  (with a target at  $v = 1$  vs.  $K - 1$  equal competitors at  $v = 0$ ), I compute the “true” choice probability  $P(\text{target})$  from the race model and then recover the best-fitting logit and probit temperature parameters by inversion.

**Figure 7**

*Recovered temperature parameters under logit (left) and probit (right) model assumptions as a function of set size  $K$  and accumulation threshold  $\theta$ .*

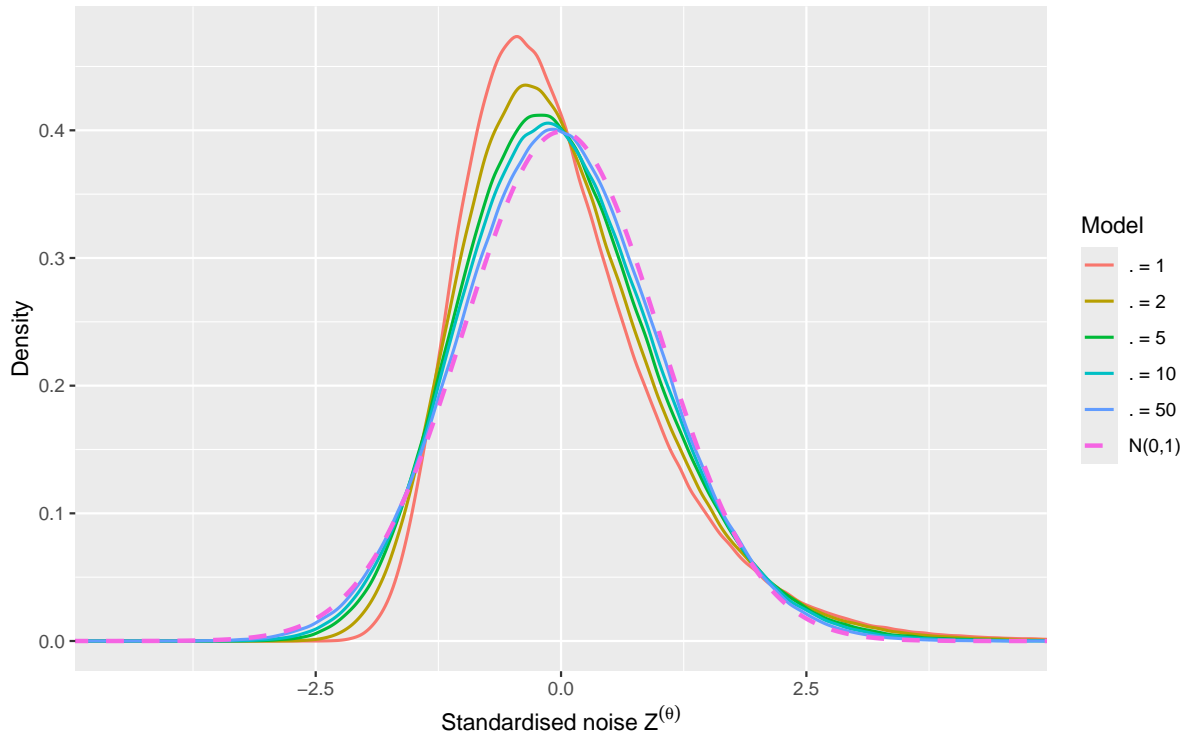


### 6.6 Study 6: Distributional Shape — Noise Skewness and Kurtosis

The log-Gamma noise distribution transitions from highly skewed (Gumbel,  $\theta = 1$ ) to symmetric (Gaussian,  $\theta \rightarrow \infty$ ). This transition in distributional shape underlies all the choice-level phenomena documented above. To make this explicit, I plot the standardised noise density for several values of  $\theta$  alongside the standard normal reference.

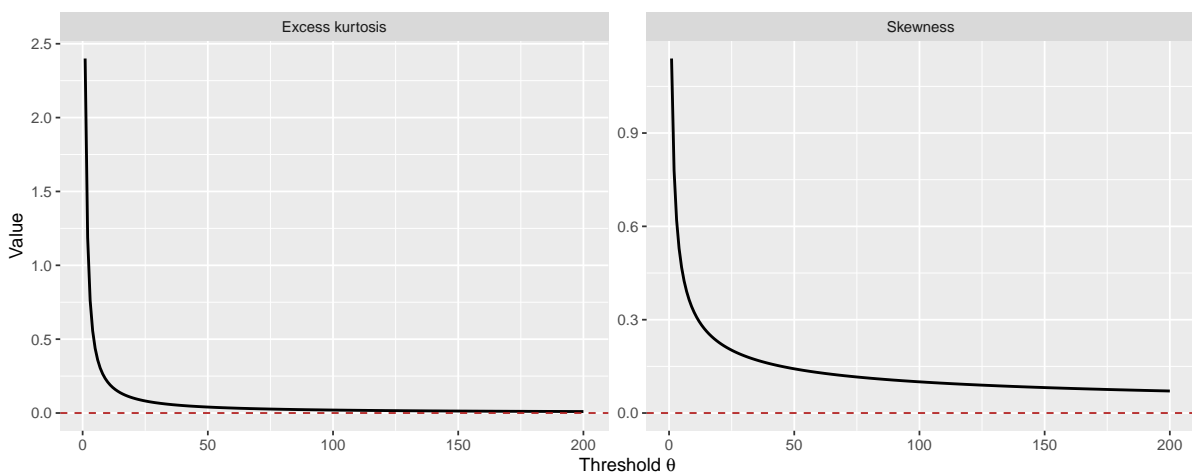
**Figure 8**

Standardised noise densities for selected values of  $\theta$ . As  $\theta$  increases, the density converges to the standard normal (black dashed).



**Figure 9**

Skewness and excess kurtosis of the standardised log-Gamma noise as a function of  $\theta$ . Both moments converge to zero (the Gaussian values) as  $\theta$  increases, with skewness decaying as  $O(\theta^{-1/2})$  and excess kurtosis as  $O(\theta^{-1})$ . These moment trajectories fully characterise the transition from Gumbel to Gaussian noise shape.

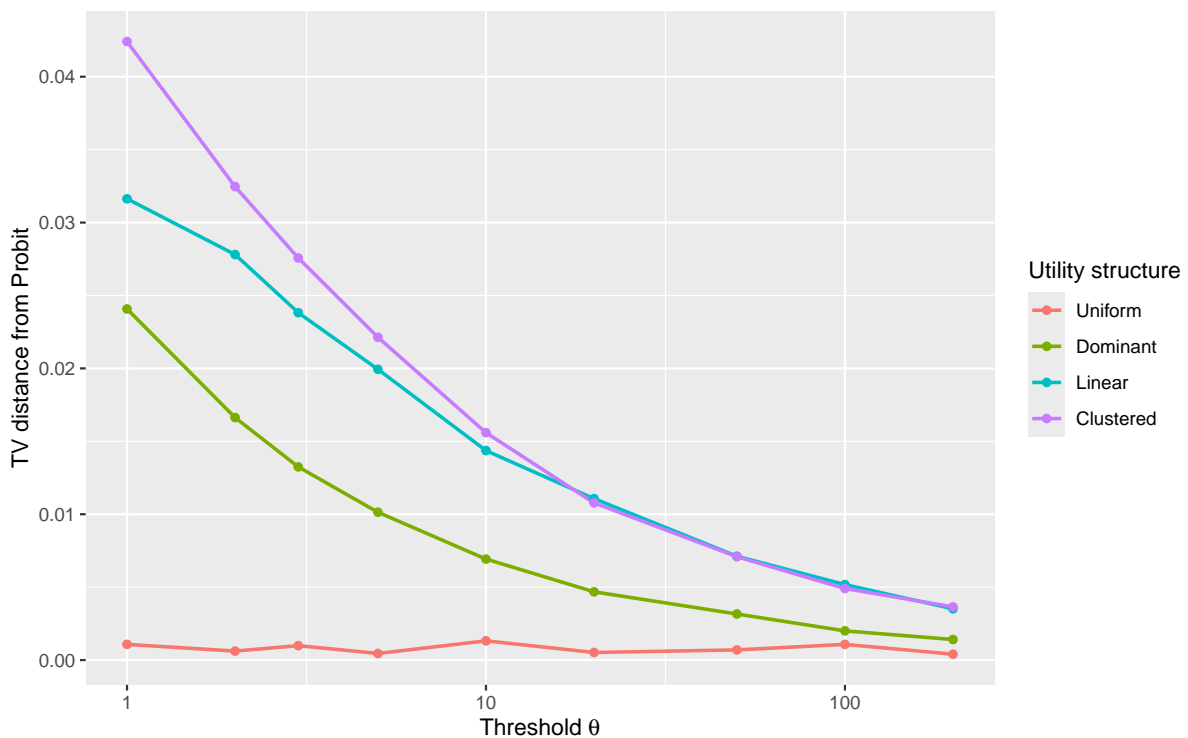


### 6.7 Study 7: Robustness Across Utility Structures

The preceding studies used specific utility vectors. To assess robustness, I examine whether the convergence pattern holds across different utility configurations that are common in psychological experiments.

**Figure 10**

*Total variation distance from Probit as a function of  $\theta$  for  $K = 5$  alternatives under four utility structures: uniform, dominant, linear, and clustered.*



### 6.8 Summary

These multinomial simulations confirm and extend the binary-case results:

1. **Convergence is gradual and universal:** Across different values of  $K$  and different utility structures, the Poisson count race converges to the Multinomial Probit reference within  $\theta \approx 50\text{--}100$  in total variation distance.
2. **Probability redistribution:** As  $\theta$  increases, the probit model concentrates more probability on the best alternative and less on inferior alternatives, reflecting the thinner tails of Gaussian noise relative to Gumbel.

3. **Set-size scaling:** The logit and probit models predict systematically different scaling of target choice probability with the number of competitors. The Poisson count race interpolates between these two patterns, connecting to the empirical findings of (Robinson et al., 2023).
4. **IIA erosion:** The Independence of Irrelevant Alternatives property, which holds exactly at  $\theta = 1$ , is progressively violated as  $\theta$  increases. This provides a process-level account of why IIA holds for logit but not probit: it is a consequence of the Gumbel noise shape, and alternative noise shapes—induced by higher accumulation thresholds—do not preserve it.
5. **Parameter invariance:** When choice data generated by the Poisson count race are fit under a logit assumption, the recovered inverse temperature drifts with set size  $K$  for all  $\theta > 1$ . Conversely, when fit under a probit assumption, the recovered noise scale remains stable for large  $\theta$  but drifts when  $\theta$  is small. This cross-over in parameter invariance provides a process-level account of the empirical findings of (Robinson et al., 2023): parameter stability across  $K$  is diagnostic of whether the effective noise distribution is closer to Gumbel or Gaussian.
6. **Noise shape transition:** The underlying mechanism is a smooth transition in the shape of the standardised noise distribution, from the skewed Gumbel ( $\theta = 1$ ) to the symmetric Gaussian ( $\theta \rightarrow \infty$ ). The skewness and kurtosis decay at known rates, providing analytic control over the approximation quality.

## 7 Extending the Race Model to Correlated Errors

The core model described thus far restricts the underlying count processes to be strictly independent, yielding the Independent Multinomial Probit in the asymptotic limit. However, the primary theoretical appeal of probit models lies in their ability to accommodate correlated error structures, thereby capturing similarity effects and systematic violations of the Independence of Irrelevant Alternatives (IIA). In this section, I demonstrate that the Poisson count race can be naturally extended to generate the full Correlated Multinomial Probit by

introducing shared evidence generators. This provides a mechanistic, process-level account for how covariance structures emerge in static choice models.

### 7.1 The Shared-Feature Mechanism

To induce correlation without violating the stationary, independent increments of the individual Poisson generators, we abandon the assumption that all events are strictly unique to a specific alternative. Instead, we define a multivariate Poisson process representing “features” or “shocks” that alternatives can share.

Let  $S \subseteq \{1, \dots, K\}$  denote a subset of alternatives. Assume the existence of independent feature generators  $Y_S(t)$ , which are Poisson processes with rates  $\lambda_S \geq 0$ . An event from  $Y_S(t)$  delivers a simultaneous unit of evidence to all alternatives contained in  $S$ . The cumulative count for alternative  $i$  is the superposition of all feature processes associated with it:

$$N_i(t) = \sum_{S:i \in S} Y_S(t)$$

Because the superposition of independent Poisson processes is itself a Poisson process,  $N_i(t)$  remains a standard Poisson process with marginal rate  $\Lambda_i = \sum_{S:i \in S} \lambda_S$ . The waiting time marginals are therefore preserved as  $T_i^{(\theta)} \sim \text{Gamma}(\theta, \Lambda_i)$ .

Crucially, for any two alternatives  $i$  and  $j$  that share at least one feature generator, their event counts covary. The covariance is determined by the sum of the rates of their shared processes:

$$\text{Cov}(N_i(t), N_j(t)) = t \sum_{S:\{i,j\} \subseteq S} \lambda_S := t\Lambda_{ij}$$

This yields a correlation between the accumulators of  $\rho_{ij} = \Lambda_{ij} / \sqrt{\Lambda_i \Lambda_j}$ .

### 7.2 The Correlated Probit Limit

To find the asymptotic joint distribution of the choice model, we apply the Multivariate Central Limit Theorem for the multivariate Poisson counting process. As  $\theta \rightarrow \infty$ , the standardized hitting times jointly converge to a Multivariate Normal distribution.

Applying the utility transformation  $U_i^{(\theta)} = \log \Lambda_i - \log G_i$ , where  $G_i = \Lambda_i T_i^{(\theta)}$  is

marginally Gamma( $\theta, 1$ ) but now correlated across alternatives, and the Multivariate Delta Method (with  $g(x) = -\log x$ ), the unstandardized random utility errors converge to a Multivariate Normal distribution whose covariance scales as  $\psi_1(\theta)\rho_{ij} \approx (1/\theta)\rho_{ij}$ .

When we apply the variance normalization from Section 3 to enforce matched discriminability, we define the standardized noise  $Z_i^{(\theta)} = (\epsilon_i^{(\theta)} - \mu_\theta)/\sigma_\theta$ . Because  $\sigma_\theta = \sqrt{\psi_1(\theta)} \approx 1/\sqrt{\theta}$  for large  $\theta$ , dividing the covariance by  $\sigma_\theta^2 = \psi_1(\theta)$  exactly recovers  $\rho_{ij}$ . The standardized noise vector  $\mathbf{Z}^{(\theta)}$  thus converges in distribution to a multivariate standard normal with the correlation matrix of the underlying counting processes:

$$\mathbf{Z}^{(\theta)} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{R})$$

where  $\mathbf{R}$  has off-diagonal elements  $\rho_{ij}$ . Consequently, the random utility vector converges to  $\mathbf{U}^{(\theta)} \xrightarrow{d} \mathcal{N}(\mathbf{v}, \beta^2 \mathbf{R})$ . This establishes the shared-feature Poisson count race as a valid generative foundation for the full Correlated Multinomial Probit.

### 7.3 Variance Normalization and Parameter Recovery

While the asymptotic proof establishes the theoretical limit, analyzing the behavior of the shared-feature model under the variance normalization introduced in Section 3 reveals critical boundary conditions of evidence accumulation.

To maintain matched discriminability across varying thresholds  $\theta$ , the systematic utilities  $v_i$  must remain constant while the noise is scaled. Because the systematic utility in the unstandardized race is  $\log(\Lambda_i)$ , the *effective total rates* must scale with  $\sigma_\theta = \sqrt{\psi_1(\theta)}$ :

$$\Lambda_i^* = \exp(v_i \sigma_\theta)$$

Consider a two-alternative scenario where options 1 and 2 share a feature generator with rate  $\lambda_{12}^*$ , and possess unique generators with rates  $\lambda_1^*$  and  $\lambda_2^*$ . To achieve a target correlation  $\rho$ , the shared rate must be:

$$\lambda_{12}^* = \rho \sqrt{\Lambda_1^* \Lambda_2^*} = \rho \exp\left(\frac{(v_1 + v_2)\sigma_\theta}{2}\right)$$

The required unique rate for alternative 2 is therefore  $\lambda_2^* = \Lambda_2^* - \lambda_{12}^*$ . For the model to

be physically coherent as a Poisson process, this unique rate must be strictly non-negative ( $\lambda_2^* \geq 0$ ).

#### 7.4 A Mechanistic Account of Substitution and Boundary Conditions

This non-negativity constraint provides a concrete process-level explanation for substitution effects. If alternative 1 is vastly superior to alternative 2 ( $v_1 \gg v_2$ ), and the correlation  $\rho$  is high, nearly all events that provide evidence for the weaker alternative are *shared* events that simultaneously provide evidence for the stronger alternative. The stronger alternative “cannibalizes” the probability mass of the weaker option because it possesses a robust unique evidence stream, while the weaker option possesses almost none.

Mathematically, ensuring  $\lambda_2^* \geq 0$  reveals a strict upper bound on the correlation the system can support for a given threshold and utility difference. Dividing by  $\Lambda_2^*$ , we find:

$$\rho_{\max} = \exp\left(-\frac{(v_1 - v_2)\sigma_\theta}{2}\right)$$

This equation demonstrates that in a physical evidence accumulation system, correlation and utility are not entirely independent parameters. A decision-maker operating at a low threshold  $\theta$  (where  $\sigma_\theta$  is large) structurally cannot sustain highly correlated representations of asymmetric options without the weaker option’s unique evidence stream collapsing below zero. However, as the decision-maker becomes more cautious ( $\theta \rightarrow \infty$ ), the scaling factor  $\sigma_\theta \rightarrow 0$ , and  $\rho_{\max} \rightarrow 1$ . The theoretical limit seamlessly supports arbitrary correlation matrices, but the generative mechanics reveal that high correlation between disparate utilities requires prolonged, thresholded accumulation.

#### 7.5 The Cost of Covariance: Degeneracy at the Logit Boundary

While independent accumulators provide a smooth, continuous bridge from the Multinomial Logit ( $\theta = 1$ ) to the Independent MNP ( $\theta \rightarrow \infty$ ), the introduction of shared features fundamentally breaks the isomorphism at the Logit boundary.

At  $\theta = 1$ , a shared event  $Y_S(t)$  reaching the single-count threshold causes simultaneous crossing for multiple alternatives. This results in a tie, violating the continuous distribution assumption necessary to derive the Luce choice rule from Extreme Value Theory and necessitating arbitrary discrete tie-breaking.

Furthermore, because incorporating a target correlation  $\rho$  requires dynamically shifting probability mass into the shared generator  $\lambda_{12}^*$ , the first two moments (variance and covariance) of the utility distribution are fixed by the target covariance structure across all  $\theta$ . Consequently, the choice probabilities exhibit Probit-like substitution patterns at all thresholds, not only in the asymptotic limit. Thus, while independent accumulators trace the transition of noise *shape* from Gumbel to Gaussian, shared accumulators sacrifice this interpolation to map the generative model onto the exact covariance structures required for general discrete choice applications.

## 8 Discussion

The present work develops a generative framework in which Multinomial Logit and Multinomial Probit arise as endpoint regimes of a single parametric family of stochastic accumulation models. By introducing a Poisson count race and a variance standardization that separates noise scale from noise shape, this paper clarifies how extreme-value and Gaussian choice behavior emerge as members of the log-Gamma random utility family, indexed by the accumulation threshold  $\theta$ . As discussed in Section 3, this bridge is distributional rather than dynamical: the variance standardization compares different accumulation systems at matched discriminability, rather than describing the behavior of a single system under threshold manipulation. The goal is not to advocate replacing existing models, but to clarify their relationship: logit and probit represent different positions within a continuum of log-Gamma noise shapes, with the accumulation threshold governing the transition between them.

The multinomial simulations demonstrate that this unification is not merely a theoretical curiosity. Thresholded accumulation induces systematic, graded violations of IIA that converge toward the dependence structure characteristic of multinomial probit. This dependence emerges endogenously from the accumulation and stopping rule, rather than being imposed by construction. The parameter invariance results further connect the framework to recent empirical findings (Robinson et al., 2023), providing a process-level account of why Gaussian-based parameters exhibit greater stability across changes in set size.

## 8.1 Relation to existing models

From a mathematical standpoint, all components of the present framework are classical: exponential races yield Luce's choice rule, Gamma waiting times arise from accumulated Poisson events, and asymptotic normality follows from the Central Limit Theorem. The contribution lies in assembling these elements into a single generative family and identifying the conditions under which its limiting behavior remains non-degenerate.

The present model should not be conflated with full sequential sampling models such as the Diffusion Decision Model ([Ratcliff, 1978](#)) or Linear Ballistic Accumulator ([Brown & Heathcote, 2005](#)). Those models jointly account for response times and accuracy via continuous accumulation with explicit drift and boundary parameters. The Poisson count race is deliberately minimal: it uses accumulation as a generative device to induce a family of random utility models, without making claims about within-trial dynamics or response time distributions.

Between the logit and probit endpoints lies a continuum of log-Gamma random utility models. These intermediate regimes are not intended as new default specifications, but they underscore that logit and probit are special cases of a broader family. Deviations from logit or probit behavior may sometimes reflect differences in accumulation thresholds rather than fundamentally different noise sources.

## 8.2 Implications for empirical modeling

The present framework offers a theoretical account of why logit-based and probit-based models may differ in parameter invariance across task structures. Models with larger effective accumulation thresholds naturally exhibit Gaussian-like behavior, which may confer greater stability across changes in the number of alternatives. At the same time, the results caution against interpreting superior empirical performance of one model class as evidence for a particular noise distribution in isolation: differences between logit and probit may reflect differences in decision criteria or commitment thresholds rather than differences in representational noise per se.

### 8.3 Limitations and extensions

The Poisson count race is intentionally simple and focuses exclusively on choice probabilities, abstracting away from response times and within-trial dynamics. Although the shared-feature extension developed in the preceding section demonstrates how correlated accumulators can generate the full Correlated Multinomial Probit, additional extensions such as time-varying rates or joint modeling of choice and response time are natural directions for future work.

The present analysis treats the accumulation threshold as fixed across trials and alternatives. Allowing threshold variability or adaptive stopping rules could further enrich the family of induced choice models and connect more directly to theories of decision caution and speed–accuracy trade-offs.

The model positions the Poisson count race as a parametric family indexed by  $(\theta, \beta)$ , but a formal identification analysis is beyond the present scope. In principle,  $\theta$  and  $\beta$  play distinct roles—shape versus scale of the noise distribution—and the shape of the psychometric function or the pattern of IIA violations could serve to identify  $\theta$  from choice data. Whether these parameters are jointly identifiable from aggregate choice frequencies alone, and under what experimental designs, remains an open question for future investigation.

Finally, the framework naturally invites comparison with other random utility specifications. Exploring whether additional classical models arise as limiting regimes under alternative accumulation rules may provide further insight into the structure of discrete choice behavior.

### 8.4 Concluding remarks

By grounding discrete choice models in a common stochastic accumulation process, the Poisson count race reframes a long-standing modeling distinction. Logit and probit emerge not as competing assumptions about utility noise, but as members of a single parametric family — the log-Gamma random utility models — indexed by the accumulation threshold  $\theta$ . This unification is algebraic and distributional: it reveals that the two canonical specifications occupy endpoint positions within a continuous family of noise shapes, rather than representing fundamentally distinct generating mechanisms. At the same time, the unification does not

imply that a single decision system can transition between regimes by adjusting its threshold alone, since the variance standardization that enables comparison across  $\theta$  values entails different effective rate structures. The framework thus clarifies the conceptual relationship between logit and probit and provides a principled basis for comparison, illustrating how process-level reasoning can illuminate the structure of static choice models.

## 9 References

- Brown, S., & Heathcote, A. (2005). A Ballistic Model of Choice Response Time. *Psychological Review*, *112*(1), 117–128. <https://doi.org/10.1037/0033-295X.112.1.117>
- Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*(1), 52–72. [https://doi.org/10.1016/0022-2496\(78\)90048-2](https://doi.org/10.1016/0022-2496(78)90048-2)
- Hausman, J. A., & Wise, D. A. (1978). A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica*, *46*(2), 403–426. <https://doi.org/10.2307/1913909>
- Kellen, D., Winiger, S., Dunn, J. C., & Singmann, H. (2021). Testing the foundations of signal detection theory in recognition memory. *Psychological Review*, *128*(6), 1022–1050. <https://doi.org/10.1037/rev0000288>
- Luce, R. D. (1959). *Individual choice behavior* (pp. xii, 153). John Wiley.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics* (p. 105).
- Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, *80*(1), 53–68. <https://doi.org/10.1037/h0033871>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>
- Robinson, M. M., DeStefano, I. C., Vul, E., & Brady, T. F. (2023). How do people build up visual memory representations from sensory evidence? Revisiting two classic models of choice. *Journal of Mathematical Psychology*, *117*, 102805. <https://doi.org/10.1016/j.jmp.2023.102805>
- Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*,

53(2), 293–315. <https://doi.org/10.1348/000711000159349>

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge University Press.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(2), 201–233.

<https://doi.org/10.1037/xlm0000732>

Yellott, J. (1977). The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2), 109–144. [https://doi.org/10.1016/0022-2496\(77\)90026-8](https://doi.org/10.1016/0022-2496(77)90026-8)