The Gumbel-Max Trick and Lehmann Family

The Gumbel-Max trick provides a method to sample from discrete distributions by transforming the sampling procedure into an optimization problem. It relies on a key property of the Gumbel distribution: we can sample from a categorical distribution $[p_1, \ldots, p_n]$ using

$$x = \arg\max_i \{\log p_i + \xi_i\}, \qquad \xi_i \overset{i.i.d.}{\sim} \texttt{Gumbel}(0, 1),$$

where the i.i.d. Gumbel noise can be generated via $\xi_i = -\log(-\log(U_i))$ with $U_i \sim \texttt{Uniform}(0, 1)$.

This note reviews the seemingly magical properties of the Gumbel distribution, which we show to be shared properties of a broader Lehmann family of distributions.

Contents

1 Distributions of Maximum and Argmax Set

Let $X_1, \ldots, X_n$ be a set of independent random variables (RVs) on $\mathbb{R}$. We are interested in studying the distribution of their maximum and the corresponding argmax set:

$$X_0 = \max\{X_1, \ldots, X_n\}, \quad J = \arg\max\{X_1, \ldots, X_n\},$$

where the argmax set J denotes the set of indices that achieve the maximum:

$$J = \{i \in \{1, \ldots, n\} \colon X_i = X_0\}.$$

The argmax set may contain multiple elements if the maximum is achieved by more than one variable. However, if the X_i are absolutely continuous, the probability of ties is zero. In this case, J contains a single element with probability one; that is, $\Pr(|J| = 1) = 1$. When this holds, we treat J as a random variable taking values in $\{1, \ldots, n\}$.

The properties of X_0 and J are most effectively characterized using the cumulative distribution functions (CDFs) of the X_i.

Recall that the CDF of a random variable X on $\mathbb{R}$ is defined as

$$F_X(x) = \Pr(X \leq x).$$

We write $X \sim F$ to denote that X has CDF F.

Lemma 1.1. *Let $X_1, \ldots, X_n$ be independent real random variables with CDFs $F_1, \ldots, F_n$. Define $X_0 = \max\{X_1, \ldots, X_n\}$ and the argmax set $J = \arg\max\{X_1, \ldots, X_n\} = \{i : X_i = X_0\}$. Then, for any $x \in \mathbb{R}$ and any $j \in \{1, \ldots, n\}$,*

1) The CDF of X_0 is

$$F_0(x) = \Pr(X_0 \leq x) = \prod_{i=1}^{n} F_i(x).$$

2) The probability that j is in the argmax set is

$$\Pr\big(j \in J\big) = \int_{-\infty}^{\infty} \prod_{i \neq j} F_i(x) \, \mathrm{d}F_j(x).$$

3) The joint law of the max and (membership of) the argmax is

$$\Pr\big(X_0 \leq x \text{ and } j \in J \big) \; = \; \int_{-\infty}^{x} \prod_{i \neq j} F_i(y) \, \mathrm{d}F_j(y).$$

Proof. (1) Since $\{X_0 \leq x\} = \bigcap_{i=1}^{n} \{X_i \leq x\}$ and the X_i are independent,

$$F_0(x) = \Pr(X_0 \leq x) = \prod_{i=1}^{n} \Pr(X_i \leq x) = \prod_{i=1}^{n} F_i(x).$$

(2) The event $\{j \in J\}$ is $\{X_j \geq X_i, \; \forall i \neq j\}$. Conditioning on X_j and using independence,

$$\Pr(j \in J) = \mathbb{E}\left[\prod_{i \neq j} \Pr(X_i \leq X_j \mid X_j)\right] = \int \prod_{i \neq j} F_i(x) \, \mathrm{d}F_j(x).$$

(3) The event $\{X_0 \leq x, \; j \in J\}$ is $\{X_j \leq x, \; X_i \leq X_j \; \forall i \neq j\}$. Conditioning on X_j,

$$\Pr(X_0 \leq x, \; j \in J) = \int_{(-\infty, x]} \prod_{i \neq j} F_i(y) \, \mathrm{d}F_j(y). \qquad \square$$

Remarks. (i) If one wants $\Pr(j$ is the *unique* argmax$) = \Pr\big(X_j > X_i \; \forall i \neq j\big)$ (strict inequalities), replace $F_i(y)$ by the left limit $F_i(y-)$ in the integrals:

$$\Pr\big(X_j > X_i \; \forall i \neq j\big) = \int \prod_{i \neq j} F_i(y-) \, \mathrm{d}F_j(y).$$

If all X_i are continuous, $F_i(y-) = F_i(y)$ a.s., ties have probability 0, and parts (2)–(3) equal the unique-argmax probabilities.

(ii) In the absolutely continuous case with densities $f_i = F_i'$, (2)–(3) reduce to

$$\Pr(j \in J) = \int_{-\infty}^{\infty} f_j(x) \prod_{i \neq j} F_i(x)\,\mathrm{d}x,$$

and

$$\Pr(X_0 \leq x,\ j \in J) = \int_{-\infty}^{x} f_j(y) \prod_{i \neq j} F_i(y)\,\mathrm{d}y.$$

2 Lehmann Family of Distributions

The general results above show that the distributions of the maximum and the argmax set rarely admit simple closed forms. An important exception arises with the *Lehmann family of distributions* (also known as the *Lehmann alternatives*), defined by the class of CDFs

$$\{F(x)^{\alpha} : \alpha > 0\},$$

where F is a fixed base distribution function.

Consider independent random variables $X_i \sim F^{\alpha_i}$ with possibly different shape parameters $\alpha_i > 0$. Then the maximum admits a simple closed form:

$$\max\{X_1, \ldots, X_n\} \ \sim\ F^{\sum_{i=1}^n \alpha_i}.$$

Moreover, if F is continuous, the distribution of the argmax set also simplifies. For each $j \in \{1, \ldots, n\}$,

$$\Pr\bigl(j \in \arg\max\{X_1, \ldots, X_n\}\bigr) \ =\ \frac{\alpha_j}{\sum_{i=1}^n \alpha_i}.$$

Moreover, in this case, the maximum $\max\{X_i\}$ and the argmax $\arg\max\{X_i\}$ are not only tractable but also *independent*. This independence is a distinctive property of the Lehmann family and highlights its analytical convenience compared with general distributions.

First, we note that positive powers of CDFs are still CDFs.

Lemma 2.1. *Let $\alpha > 0$, and $G(x) = F(x)^{\alpha}$, then F is a CDF iff G is a CDF.*

Remark For positive integer n, $F(x)^n$ is the CDF of $X_0 = \max(X_1, \ldots, X_n)$, if $X_i \overset{i.i.d.}{\sim} F$.

Proof. A function $F\colon \mathbb{R} \to [0,1]$ is a CDF of some random variable iff it is a *cadlag* function (i.e., *non-decreasing*, and *right-continuous*), and

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1. \tag{1}$$

It is straightforward to verify that F satisfies these properties iff G does so. □

Definition 2.2 (Lehmann family (power transformations of CDFs))**.** *Let F be a cumulative distribution function (CDF) on $\mathbb{R}$, and let $\alpha > 0$. Define the* Lehmann family *generated by F as*

$$F_{\alpha}(x) \coloneqq F(x)^{\alpha}, \qquad x \in \mathbb{R},\ \alpha > 0.$$

Main Result We now provide explicit formulas for the distributions of the maximum and the argmax in the Lehmann family of distributions.

Theorem 2.3 (Max and Argmax of Lehmann.). *Let $X_1, \ldots, X_n$ be independent random variables on $\mathbb{R}$, where the CDF F_i of X_i satisfies*

$$F_i(x) = F(x)^{\alpha_i},$$

for some positive constants $\{\alpha_i > 0\}$ and a base CDF F. Assume F is continuous. Then:

1) The CDF of the maximum $X_0 = \max\{X_1, \ldots, X_n\}$ is

$$F_{X_0}(x) = F(x)^{\sum_{i=1}^n \alpha_i}.$$

2) The argmax set $J = \arg\max\{X_1, \ldots, X_n\}$ satisfies

$$\Pr(j \in J) = \frac{\alpha_j}{\sum_{i=1}^n \alpha_i}.$$

3) The joint distribution of the maximum X_0 and the argmax J satisfies:

$$\Pr(\{X_0 \le x \text{ and } j \in J\}) = \frac{\alpha_j}{\sum_{i=1}^n \alpha_i} \cdot F(x)^{\sum_{i=1}^n \alpha_i}, \qquad \forall x, j.$$

In fact, maximum X_0 and argmax set J are independent:

$$\Pr(\{X_0 \le x \text{ and } j \in J\}) = \Pr(X_0 \le x) \cdot \Pr(j \in J).$$

Proof. 1) Since the variables are independent,

$$F_{X_0}(x) = \prod_i F_i(x) = \prod_i F(x)^{\alpha_i} = F(x)^{\sum_i \alpha_i}.$$

2) From Lemma 1.1, we have

$$\begin{aligned}
\Pr(j \in J) &= \int \prod_{i \ne j} F(x)^{\alpha_i} \, \mathrm{d}F^{\alpha_j}(x) \\
&= \alpha_j \int F(x)^{\sum_{i \ne j} \alpha_i + \alpha_j - 1} \, \mathrm{d}F(x) \qquad // \mathrm{d}F^{\alpha_j} = \alpha_j F^{\alpha_j - 1} \mathrm{d}F \\
&= \alpha_j \int_I t^{\sum_i \alpha_i - 1} \, \mathrm{d}t,
\end{aligned}$$

where we use the substitution $t = F(x)$ and $I = \{F(x) \colon x \in \mathbb{R}\}$.

Since F is continuous, $I = [0, 1]$. Therefore,

$$\Pr(j \in J) = \alpha_j \int_0^1 t^{\sum_i \alpha_i - 1} \, \mathrm{d}t = \frac{\alpha_j}{\sum_i \alpha_i}.$$

3) Again using Lemma 1.1, we have

$$\begin{aligned}
&\Pr(\{X_0 \le x \text{ and } j \in J\}) \\
&= \int_{-\infty}^{x} \prod_{i \ne j} F_i(y)\,\mathrm{d}F_j(y) \\
&= \alpha_j \int_{-\infty}^{x} F(y)^{\sum_{i=1}^n \alpha_i - 1}\,\mathrm{d}F(y) \\
&= \frac{\alpha_j}{\sum_{i=1}^n \alpha_i} \left(F(x)^{\sum_{i=1}^n \alpha_i} - F(-\infty)^{\sum_{i=1}^n \alpha_i} \right) \\
&= \frac{\alpha_j}{\sum_{i=1}^n \alpha_i} \cdot F(x)^{\sum_{i=1}^n \alpha_i}.
\end{aligned}$$

On the other hand, recall that $\Pr(j = \arg\max\{X_i\}) = \frac{\alpha_j}{\sum_{i=1}^n \alpha_i}$, and $\Pr(X_0 \le x) = F(x)^{\sum_{i=1}^n \alpha_i}$. We get $\Pr(\{X_0 \le x \text{ and } j \in J\}) = \Pr(X_0 \le x) \cdot \Pr(j = J)$.

□

Sampling from the Lehmann Family We consider how to generate random variables with CDF $F(x)^\alpha$, given access to a sampler for $F(x)$.

Recall a basic result from *inverse transform sampling*: If F is continuous and strictly increasing, then $U = F(X)$ follows $\texttt{Uniform}([0,1])$, and $X = F^{-1}(U)$ follows F when $U \sim \texttt{Uniform}([0,1])$. If F is not invertible, we have the following generalization.

1) $F^{-1}(U)$ follows F when $U \sim \texttt{Uniform}([0,1])$, where F^{-1} denotes the generalized inverse of F:

$$F^{-1}(u) = \inf\{x \in \mathbb{R} \colon F(x) \ge u\}, \quad \forall u \in [0,1],$$

also known as the quantile function.

2) $U = \hat{F}(X)$ follows $\texttt{Uniform}([0,1])$ if $\hat{F}$ is a randomized function defined by

$$\hat{F}(x) = \begin{cases} F(x) & \text{if } F \text{ is continuous at } x, \\ Z & \text{where } Z \sim \texttt{Uniform}([F_-(x), F_+(x)]) \text{ otherwise,} \end{cases}$$

with $F_\pm(x) = \lim_{y \to x^\pm} F(y)$ denoting the left and right limits of F at x.

Lemma 2.4. *A random variable Y with CDF $F(x)^\alpha$ can be constructed in either of the following ways:*

1) $Y = F^{-1}(U^{1/\alpha})$ with $U \sim \texttt{Uniform}(0,1)$.

2) $Y = F^{-1}(\hat{F}(X)^{1/\alpha})$ with $X \sim F$.

Proof. We want to solve $F^\alpha(Y) = U = \hat{F}(X)$. It gives $Y = F^{-1}(U^{1/\alpha}) = F^{-1}(\hat{F}(X)^{1/\alpha})$. □

3 Gumbel-Max Trick

Definition 3.1 (Gumbel Distribution)**.** *A random variable X follows a Gumbel distribution* $\texttt{Gumbel}(\mu, \sigma)$ *if its CDF and density function are, respectively,*

$$F_{\mu,\sigma}(x) = \exp(-\exp(-\frac{x-\mu}{\sigma})),$$
$$p_{\mu,\sigma}(x) = F_{\mu,\sigma}(x) \cdot \exp(-\frac{x-\mu}{\sigma}) \cdot \frac{1}{\sigma}.$$

An key consequence of the definition is that

$$F_{\mu,\sigma}(x) = F_{0,\sigma}(x)^{\exp(\mu/\sigma)}, \qquad F_{0,\sigma}(x) = \exp(-\exp(-x/\sigma)), \qquad \forall \mu,$$

where $F_{0,\sigma}(x)$ is the CDF of $\texttt{Gumbel}(0, \sigma)$. Hence, Gumbel distributions are a Lehmann family. This leads to the following key result as a direct consequence of Theorem 2.3.

Theorem 3.2 (The Gumbel-Max Trick)**.** *Let $X_i \sim \texttt{Gumbel}(\mu_i, \sigma)$, $i = 1, \ldots, d$ are independent Gumbel random variables.*

1) The maximum $X_0 \overset{def}{=} \max_i\{X_i : i = 1, \ldots, n\}$ follows $\texttt{Gumbel}(\mu_0, \sigma)$, where

$$\mu_0 = \Phi_\sigma(\{\mu_i\}) \overset{def}{=} \sigma \log \sum_{i=1}^{n} \exp\left(\frac{\mu_i}{\sigma}\right).$$

2) With probability one the argmax set $J = \arg\max\{X_i\}$ only includes one element (i.e., no tied maximum), and

$$\Pr(j = J) = \frac{\exp(\mu_j/\sigma)}{\sum_{i=1}^{n} \exp(\mu_i/\sigma)}.$$

3) The maximum $X_0 = \max\{X_i\}$ and argmax $\arg\max\{X_i\}$ are independent:

$$\Pr(\{X_i \le x, \text{ and } j = J\}) = \Pr(X_0 \le x) \times \Pr(j = J), \qquad \forall x, j.$$

Remark If $X_i \sim \texttt{Gumbel}(\mu_i, \sigma_i)$ with different scaling factor σ_i, then the nice properties above no longer hold.

Proof. Just apply Theorem 2.3 with $F(x) = F_{0,\sigma}(x)$ and $\alpha_i = \exp(\mu_i/\sigma)$.

But we also provide the direct derivations here. Let $F_i(x)$ be the distribution function of X_i.

$$\begin{aligned} F_0(x) = \prod_i F_i(x) &= \exp\big[-\sum_i \exp(-\frac{x-\mu_i}{\sigma})\big] \\ &= \exp\big[-\exp(-\frac{x}{\sigma})(\sum_i \exp(\frac{\mu_i}{\sigma}))\big] \\ &= \exp\big[-\exp(-\frac{x-\Phi_\sigma(\{\mu_i\})}{\sigma})\big]. \end{aligned}$$

Assume $\alpha = 1$ without loss of generality. We have

$$
\begin{aligned}
\Pr(1 \in \arg\max\{X_i\}) &= \int_{x_1} \prod_{i>1} F_i(x) \mathrm{d}F_1(x) \\
&= \int_{x_1} (\prod_{i=1} F_i(x_1)) \exp(-\frac{x_1 - \mu_1}{\sigma}) \frac{1}{\sigma} \mathrm{d}x \\
&= \int_{x_1} (F_0(x)) \exp(-\frac{x_1 - \mu_1}{\sigma}) \frac{1}{\sigma} \mathrm{d}x_1 \qquad // X_0 = \max_i X_i \text{ by Lemma 3.2} \\
&= \int_{x_1} (p_{X_0}(x_1)) \exp(\mu_1/\sigma)/(\sum_i \exp(\mu_i/\sigma)) dx_1 \\
&= \exp(\mu_1/\sigma)/(\sum_i \exp(\mu_i/\sigma))
\end{aligned}
$$

Since X_i have absolutely continuous density functions, the probability of having tied maximum is zero, that is, $\Pr(|\arg\max\{X_i\}|) = 0$. Hence we can write $\Pr(j \in \arg\max\{X_i\}) = \Pr(j = \arg\max\{X_i\})$ in all cases. □

Remark Basic properties of $\texttt{Gumbel}(\mu, \sigma)$:

1. Its mode is μ; its median is $\mu - \log(\log 2)\sigma$.
2. The mean is $\mathbb{E}(X) = \mu + \gamma\sigma$, where $\gamma \approx 0.5772157$ is the Euler-Mascheroni constant.
3. Its standard deviation is $\sigma\pi/\sqrt{6}$.
4. If $U \sim \texttt{Uniform}([0,1])$, then $X = -\log(-\log(U))$ follows $\texttt{Gumbel}(0,1)$.

Remark Using the mean/median/mode formulas of Gumbel distributions, we have

$$
\begin{aligned}
\Phi_\sigma(\{\mu_i\}) &= \mathbb{E}[\max\{X_1, \ldots, X_n\}] - \gamma\sigma \\
&= \mathrm{Median}[\max\{X_1, \ldots, X_n\}] + \log(\log(2))\sigma \\
&= \mathrm{Mode}[\max\{X_1, \ldots, X_n\}].
\end{aligned}
$$

These can be used to construct stochastic estimation of $\Phi_\sigma(\{\mu_i\})$.

Question Since Theorem 2.3 works for a broader family of random variables, can we find other interesting families beyond Gumbel? Unfortunately, the answer is somewhat negative, since all families with continuous F under Theorem 2.3 are essentially equivalent in the following sense.

Theorem 3.3. *Let F and G be two CDF on $\mathbb{R}$ and $X_1, \ldots, X_n$ are independent random variables with CDF $F(x)^{\alpha_i}$ for X_i, where $\alpha_i > 0$. Define a non-decreasing mapping $T(x) = G^{-1}(\hat{F}(x))$. Then $Y_i = T(X_i)$ yields CDF $G(x)_i^\alpha$.*

Proof. Because $U_i \overset{\text{def}}{=} \hat{F}(X_i)^{\alpha_i}$ follows $\texttt{Uniform}([0,1])$, we can find a Y_i with CDF G^{α_i} by solving $U_i = G(Y_i)^{\alpha_i}$. This yields $U_i = F(X_i)^{\alpha_i} = G(Y_i)^{\alpha_i}$ and hence $Y_i = G^{-1}(F(X_i))$. □

Remark Because T is a non-decreasing mapping (with fixed random seed of $\hat{F}$ if F is discontinuous), they have the same properties in terms of maximum and argmax distributions. In particular, we have we have $\max(Y_1, \dots Y_n) = T(\max(X_1, \dots X_n))$ and $\arg\max(X_1, \dots X_n) \subset \arg\max(Y_1, \dots, Y_n)$. Moreover, when F and G are continuous, the mapping T is deterministic and invertible, and we have $\arg\max(X_1, \dots X_n) = \arg\max(Y_1, \dots, Y_n)$.

Remark Gumbel distribution is special in that it makes α a location parameter, that is, we have $X + \log\alpha \sim F(x)^\alpha$ if $X \sim F$ with $F(x) = \exp(-\exp(x))$. This is because by Lemma 2.4 we can draw from F^α with $Y = F^{-1}(F(X)^{1/\alpha}) = \log(-\log(\exp(-\exp(X))^\alpha)) = X + \log\alpha$.

Remark Let $X_1, \dots, X_n$ be independent with $X_i \sim \texttt{Exp}(\alpha_i)$ whose CDF is $F_i(x) = 1 - \exp(-\alpha_i x)$ for $x \in [0, +\infty)$. It is well known that

$$\min_i\{X_1, \dots, X_n\} \sim \texttt{Exp}(\alpha_1 + \dots + \alpha_n),$$

and

$$\Pr(j = \arg\min\{X_1, \dots, X_n\}) = \frac{\alpha_i}{\alpha_1 + \dots + \alpha_n}.$$

This directly follows Theorem 2.3 by noting that $Y_i = -X_i$ has a CDF of $F_i(x) = \exp(\alpha_i x)$ for $x \in (-\infty, 0]$.

Note that $-Y \sim \texttt{Exp}(\alpha_i)$ if $Y = \log U/\alpha = -\exp(X)/\alpha$ with $X \sim \texttt{Gumbel}(0,1)$ and $U \sim \texttt{Uniform}([0,1])$.

4 Sampling From Truncated and Posterior Gumbel Distributions

Definition 4.1. *For a random variable X on $\mathbb{R}$ with CDF $F(x)$, the truncated distribution of X on an intervals $(a, b]$ is the distribution of X conditioned on $X \in (a, b]$. It can be realized by rejection sampling*

$$Y = \begin{cases} X & \text{if } X \in (a,b] \\ \texttt{Reject} & \text{if } X \notin (a,b]. \end{cases}$$

The CDF of the truncated random variable Y is

$$F_Y(x) = \Pr(X \le x \mid X \in (a,b]) = \frac{F(x) - F(a)}{F(a) - F(b)}.$$

Using inverse transform sampling, one can alternatively generate Y via

$$Y = F^{-1}(U), \qquad \text{with} \qquad U \sim \texttt{Uniform}([F(a),\ F(b)]). \tag{2}$$

For upper truncated Gumbel distributions, we have an elegant result related to log-sum-exp function.

Theorem 4.2. *The $\texttt{Gumbel}(\mu, \sigma)$ distribution truncated at $(-\infty, b]$ can be realized by*

$$Y = -\sigma\log(\exp(-X/\sigma) + \exp(-b/\sigma)), \qquad \text{with} \qquad X \sim \texttt{Gumbel}(\mu, \sigma).$$

Remark If $\sigma \to +\infty$, we have $Y \to \min(X, b)$. Hence, Y is a *soft* minimum of X and b.

Proof. Let $F(x) = \exp(-\exp(-(x-\mu)/\sigma))$ be the CDF of $\texttt{Gumbel}(\mu, \sigma)$. Using (2), the truncated distribution on $(-\infty, b]$ can be sampled by

$$Y = F^{-1}(F(b)U) = -\sigma \log(-\log(F(b)U)) + \mu = -\sigma \log(-\log U - \log F(b)) + \mu,$$

where $U \sim \texttt{Uniform}([0,1])$ and hence $F(b)U \sim \texttt{Uniform}([0, F(b)])$.

Note $\log F(b) = -\exp(-(b-\mu)/\sigma)$, and $\log U = -\exp(-(X-\mu)/\sigma)$ for $X = -\sigma \log(-\log U) + \mu \sim \texttt{Gumbel}(0,1)$. We have

$$\begin{aligned} Y &= -\sigma \log(\exp(-(X-\mu)/\sigma) + \exp(-(b-\mu)/\sigma)) + \mu \\ &= -\sigma \log(\exp(-X/\sigma) + \exp(-b/\sigma)). \end{aligned}$$

□

Posterior Gumbel Sampling Assume $X_i \sim \texttt{Gumbel}(\mu_i, \sigma)$ are independent random variables. How to draw a posterior sample of $\{X_1, \ldots, X_n\}$ conditioned on $j = \arg\max(X_1, \ldots, X_n)$?

Because the distribution of $X_0 \stackrel{\text{def}}{=} \max(X_1, \ldots, X_n)$ is independent with $\arg\max(X_1, \ldots, X_n)$ by Theorem 2.3. We can first draw X_0, which follows $\texttt{Gumbel}(\mu_0, \sigma)$ with $\mu_0 = \sigma \log \sum_i \exp(\mu_i/\sigma)$, and then draw X_i conditioned on $X_i \leq X_0$, which follows $\texttt{Gumbel}(\mu, \sigma)$ truncated on $(0, X_0]$ (see Theorem 4.2):

$$Y_i = \begin{cases} X_i + (\mu_0 - \mu_i) & \text{if } i = j \\ -\sigma \log(\exp(-X_i/\sigma) + \exp(-Y_j/\sigma)) & \text{if } i \neq j, \end{cases}$$

where X_i are drawn independently from $\texttt{Gumbel}(\mu_i, \sigma)$.