Discrete choice models with taste shocks

Richard Foltyn

University of Glasgow

March 26, 2022

In recent years, it has become common practice to augment macroeconomic models that include discrete choices with taste shocks to smooth out kinks and non-differentiabilities. This note derives the expressions for expected utility and choice probabilities for binary and multinomial choices models with taste shocks. Derivatives of expected values are provided which are required to for numerical solution methods. Additionally, common pitfalls arising in numerical implementations are discussed.

Contents

1 Binary choice model

1.1 Expected value without distributional assumptions

We consider a discrete choice problem between two alternatives indexed by $i \in \{1, 2\}$. Let $\boldsymbol{x}$ be a household's state vector, and $V_i(\boldsymbol{x})$ the value function when choosing alternative i at state $\boldsymbol{x}$. We assume there is an additive alternative-specific draw ϵ_i which is i.i.d.

and independent of $\boldsymbol{x}$ such that the household's utility after the realization of ϵ_i is

$$V(\boldsymbol{x}) = \max_i \left\{ V_i(\boldsymbol{x}) + \epsilon_i \,\middle|\, i \in \{1,2\} \right\}.$$

We are interested in computing two objects: the expected value function

$$\mathbf{E}V(\boldsymbol{x}) = \mathbf{E}\left[\max\left\{ V_1(\boldsymbol{x}) + \epsilon_1,\ V_2(\boldsymbol{x}) + \epsilon_2 \right\} \right], \tag{1}$$

and the probability that alternative i will be chosen,

$$\pi_i = \Pr\left(V_i(\boldsymbol{x}) + \epsilon_i \geq V_{-i}(\boldsymbol{x}) + \epsilon_{-i} \right) \tag{2}$$

where we adopt the usual notation that when one choice is indexed by i, the other remaining alternative is denoted by $-i$.[1] In the remainder of the document, we drop the explicit dependence on $\boldsymbol{x}$ and write V_i instead of $V_i(\boldsymbol{x})$.

The expectation in (1) can be written as follows:

$$\begin{aligned}
\mathbf{E}V(\boldsymbol{x}) &= \mathbf{E}\left[\max\left\{ V_1 + \epsilon_1,\ V_2 + \epsilon_2 \right\} \right] \\
&= \mathbf{E}\left[V_1 + \epsilon_2 \,\middle|\, V_1 + \epsilon_1 \geq V_2 + \epsilon_2 \right] \times \Pr\left(V_1 + \epsilon_1 \geq V_2 + \epsilon_2 \right) \\
&\quad + \mathbf{E}\left[V_2 + \epsilon_2 \,\middle|\, V_2 + \epsilon_2 > V_1 + \epsilon_2 \right] \times \Pr\left(V_2 + \epsilon_2 > V_1 + \epsilon_1 \right)
\end{aligned} \tag{3}$$

Let $\Delta_i \equiv V_i - V_{-i}$ be the difference in value functions before the shock realization. Denote the CDF of ϵ_i by $F_\epsilon(\bullet \mid \mu_i, \sigma)$ and its PDF by $f_\epsilon(\bullet \mid \mu_i, \sigma)$, and assume that its support is $[\underline{\epsilon}, \overline{\epsilon}]$, where either of them is permitted to be infinite. Note that we assume a common scale parameter σ, but allow for alternative-specific location parameters μ_i.

First, consider the probability of choosing alternative i: we have that

$$\begin{aligned}
\pi_i &= \Pr\left(V_i + \epsilon_i \geq V_{-i} + \epsilon_{-i} \right) = \Pr\left(\Delta_i + \epsilon_i \geq \epsilon_{-i} \right) \\
&= \int_{\underline{\epsilon}}^{\overline{\epsilon}} \left[\int_{\underline{\epsilon}}^{\Delta_i + \epsilon_i} f_\epsilon(\epsilon_{-i} \mid \mu_{-i}, \sigma)\, d\epsilon_{-i} \right] f_\epsilon(\epsilon_i \mid \mu_i, \sigma)\, d\epsilon_i \\
&= \int_{\underline{\epsilon}}^{\overline{\epsilon}} F_\epsilon(\Delta_i + \epsilon_i \mid \mu_{-i}, \sigma)\, f_\epsilon(\epsilon_i \mid \mu_i, \sigma)\, d\epsilon_i
\end{aligned} \tag{4}$$

[1] We assume that the ϵ are continuous random variables, so it makes no difference whether we use a weak or strict inequality sign in (2).

Next, the conditional expectation in (3) can be written as

$$
\begin{aligned}
&\mathbf{E}\left[V_i + \epsilon_i \,\middle|\, V_i + \epsilon_i \geq V_{-i} + \epsilon_{-i} \right] \\
&\quad = \mathbf{E}\left[V_i + \epsilon_i \,\middle|\, \Delta_i + \epsilon_i \geq \epsilon_{-i} \right] \\
&\quad = \int_{\underline{\epsilon}}^{\overline{\epsilon}} \int_{\underline{\epsilon}}^{\Delta_i+\epsilon_i} (V_i + \epsilon_i) \frac{f_\epsilon\left(\epsilon_{-i} \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right)}{\pi_i} d\epsilon_{-i} d\epsilon_i \\
&\quad = \frac{1}{\pi_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} (V_i + \epsilon_i) \left[\int_{\underline{\epsilon}}^{\Delta_i+\epsilon_i} f_\epsilon\left(\epsilon_{-i} \,|\, \mu_{-i}, \sigma\right) d\epsilon_{-i} \right] f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) d\epsilon_i \\
&\quad = \frac{1}{\pi_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} (V_i + \epsilon_i) F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) d\epsilon_i \\
&\quad = V_i + \frac{1}{\pi_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_i F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) d\epsilon_i
\end{aligned}
$$

Therefore, we find that

$$
\begin{aligned}
\mathbf{E}\left[V_i + \epsilon_i \,\middle|\, V_i + \epsilon_i \geq V_{-i} + \epsilon_{-i} \right] \times \Pr\left(V_i + \epsilon_i \geq V_{-i} + \epsilon_{-i} \right)& \\
= \pi_i V_i + \int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_i F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) d\epsilon_i&
\end{aligned}
$$

and the "upper envelope" in (1) is consequently given by

$$
\begin{aligned}
\mathbf{E}V = \pi_1 V_1 + \pi_2 V_2 + \int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_1 F_\epsilon\left(V_1 - V_2 + \epsilon_1 \,|\, \mu_2, \sigma\right) f_\epsilon\left(\epsilon_1 \,|\, \mu_1, \sigma\right) d\epsilon_1& \\
+ \int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_2 F_\epsilon\left(V_2 - V_1 + \epsilon_2 \,|\, \mu_1, \sigma\right) f_\epsilon\left(\epsilon_2 \,|\, \mu_2, \sigma\right) d\epsilon_2&
\end{aligned}
$$

1.2 Type-I extreme value distribution

We now make specific assumptions on the taste shock distribution and let ϵ_i be draws from the type-I extreme value (or Gumbel) distribution with location parameter $\mu \in (-\infty, \infty)$ and scale parameter $\sigma > 0$. Note that the Gumbel distribution has support on $\mathbb{R}$. The CDF F_ϵ and PDF f_ϵ are given by

$$
F_\epsilon\left(x \,|\, \mu, \sigma\right) = \exp\Big\{ -\exp\Big(-(x-\mu)/\sigma \Big) \Big\} \tag{5}
$$

$$
f_\epsilon\left(x \,|\, \mu, \sigma\right) = \frac{1}{\sigma} \exp\left\{ -\left(\frac{x-\mu}{\sigma} \right) - \exp\Big(-(x-\mu)/\sigma \Big) \right\} \tag{6}
$$

We are interested in evaluating the expression $F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right)$ which in the case of the Gumbel distribution is

$$
\begin{aligned}
&F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) \\
&\quad = \exp\left\{-e^{-(\Delta_i+\epsilon_i-\mu_{-i})/\sigma}\right\} \times \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i}{\sigma}\right) - e^{-(\epsilon_i-\mu_i)/\sigma}\right\} \\
&\quad = \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i}{\sigma}\right) - e^{-(\Delta_i+\epsilon_i-\mu_{-i})/\sigma} - e^{-(\epsilon_i-\mu_i)/\sigma}\right\} \\
&\quad = \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i}{\sigma}\right) - e^{-(\epsilon_i-\mu_i)/\sigma}\left[1 + e^{-(\Delta_i+\Delta\mu_i)/\sigma}\right]\right\}
\end{aligned}
$$

where we define $\Delta\mu_i \equiv \mu_i - \mu_{-i}$. Let α_i denote the term

$$\alpha_i \equiv 1 + e^{-(\Delta_i+\Delta\mu_i)/\sigma}$$

Then, continuing with the last line above, we have

$$
\begin{aligned}
&F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) \\
&\quad = \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i}{\sigma}\right) - e^{-(\epsilon_i-\mu_i)/\sigma}\alpha_i\right\} \\
&\quad = \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i}{\sigma}\right) - e^{-(\epsilon_i-\mu_i)/\sigma}e^{\log\alpha_i}\right\} \\
&\quad = \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i-\sigma\log\alpha_i}{\sigma}\right) - \log\alpha_i - e^{-(\epsilon_i-\mu_i-\sigma\log\alpha_i)/\sigma}\right\} \\
&\quad = \frac{1}{\alpha_i} \times \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\widehat{\mu}_i}{\sigma}\right) - e^{-(\epsilon_i-\widehat{\mu}_i)/\sigma}\right\} \\
&\quad = \frac{1}{\alpha_i} \times f_\epsilon\left(\epsilon_i \,|\, \widehat{\mu}_i, \sigma\right)
\end{aligned}
$$

where we define

$$\widehat{\mu}_i \equiv \mu_i + \sigma\log\alpha_i = \mu_i + \sigma\log\left(1 + e^{-(\Delta_i+\Delta\mu_i)/\sigma}\right)$$

and $f_\epsilon\left(\bullet \,|\, \widehat{\mu}_i, \sigma\right)$ is the Gumbel PDF with location parameter $\widehat{\mu}_i$ and scale parameter σ.

We are now ready to revisit the generic ex-ante probability of choosing alternative i in (4), applying the Gumbel-specific expression derived above:

$$
\begin{aligned}
\pi_i = \int_{\underline{\epsilon}}^{\overline{\epsilon}} F_\epsilon\left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon\left(\epsilon_i \,|\, \mu_i, \sigma\right) d\epsilon_i = \frac{1}{\alpha_i}\int_{\underline{\epsilon}}^{\overline{\epsilon}} f_\epsilon\left(\epsilon_i \,|\, \widehat{\mu}_i, \sigma\right) d\epsilon_i &= \frac{1}{\alpha_i} \\
&= \frac{1}{1 + e^{-\left((V_i+\mu_i)-(V_{-i}+\mu_{-i})\right)/\sigma}} \\
&= \frac{e^{(V_i+\mu_i)/\sigma}}{e^{(V_i+\mu_i)/\sigma} + e^{(V_{-i}+\mu_{-i})/\sigma}}
\end{aligned}
$$

We see that the probability of choosing alternative i is the usual expression from the Logit model. Using the result $\pi_i = \alpha_i^{-1}$, we can also rewrite the location parameter $\widehat{\mu}_i$ as

$$\widehat{\mu}_i = \mu_i - \sigma \log \pi_i$$

Turning to the expected value $\mathbf{E}V$, we need to evaluate the expected value of the "modified" Gumbel distribution

$$\int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_i F_\epsilon \left(\Delta_i + \epsilon_i \,|\, \mu_{-i}, \sigma\right) f_\epsilon \left(\epsilon_i \,|\, \mu_i, \sigma\right) d\epsilon_i$$
$$= \frac{1}{\alpha_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_i f_\epsilon(\epsilon_i \,|\, \widehat{\mu}_i, \sigma) d\epsilon_i = \pi_i \Big[\widehat{\mu}_i + \sigma\gamma\Big] = \pi_i \Big[\mu_i - \sigma \log \pi_i + \sigma\gamma\Big] \quad (7)$$

where $\gamma \approx 0.57721$ is the Euler-Mascheroni constant. Consequently,

$$\mathbf{E}V = \pi_1 V_1 + \pi_2 V_2 + \pi_1 \Big(\mu_1 - \sigma \log \pi_1 + \sigma\gamma\Big) + \pi_2 \Big(\mu_2 - \sigma \log \pi_2 + \sigma\gamma\Big)$$

Note that

$$\begin{aligned} \pi_i \sigma \log \pi_i &= \pi_i \sigma \Big[\log \left(e^{(V_i+\mu_i)/\sigma}\right) - \log \left(e^{(V_i+\mu_i)/\sigma} + e^{(V_{-i}+\mu_{-i})/\sigma}\right)\Big] \\ &= \pi_i V_i + \pi \mu_i - \pi_i \sigma \log \left(e^{(V_i+\mu_i)/\sigma} + e^{(V_{-i}+\mu_{-i})/\sigma}\right) \end{aligned}$$

Thus the above expression can be simplified to

$$\begin{aligned} \mathbf{E}V &= \sigma \log \left(e^{(V_1+\mu_1)/\sigma} + e^{(V_2+\mu_2)/\sigma}\right) + \sigma\gamma \\ &= \sigma \log \left(e^{\mu_2/\sigma} \left(e^{(V_1+\Delta\mu_1)/\sigma} + e^{V_2/\sigma}\right)\right) + \sigma\gamma \\ &= \sigma \log \left(e^{(V_1+\Delta\mu_1)/\sigma} + e^{V_2/\sigma}\right) + \mu_2 + \sigma\gamma \end{aligned}$$

where $\Delta\mu_1 = \mu_1 - \mu_2$ as before, and we also use the fact that $\pi_1 = 1 - \pi_2$. Note that if $\mu_1 = \mu_2$ and thus $\Delta\mu_1 = 0$, this expression simplifies to the more familiar

$$\mathbf{E}V = \sigma \log \left(e^{V_1/\sigma} + e^{V_2/\sigma}\right) + \mu + \sigma\gamma$$

2 Multinomial choice model

2.1 Expected value without distributional assumptions

Consider now the generalization to $n \geq 2$ alternatives. Then the value function is given by

$$V(\boldsymbol{x}) = \max_i \Big\{ V_i(\boldsymbol{x}) + \epsilon_i \,\Big|\, i \in \{1, \ldots, n\} \Big\}$$

and its expectation over taste shock realizations is

$$\mathbf{E}V(\boldsymbol{x}) = \mathbf{E}\left[\max\Big\{V_1(\boldsymbol{x}) + \epsilon_1,\ \ldots,\ V_n(\boldsymbol{x}) + \epsilon_n\Big\}\right]$$

with the the probability that alternative i will be chosen given by

$$\pi_i = \Pr\left(V_i(\boldsymbol{x}) + \epsilon_i \geq \max_{j \neq i} \{V_j + \epsilon_j\} \right)$$

The expectation above can be written as follows:

$$\begin{aligned}
\mathbf{E}V(\boldsymbol{x}) &= \mathbf{E}\left[\max\left\{ V_1 + \epsilon_1, \ldots, V_n + \epsilon_n \right\} \right] \\
&= \sum_i \mathbf{E}\left[V_i + \epsilon_i \,\middle|\, V_i + \epsilon_i \geq \max_{j \neq i} \{V_j + \epsilon_j\} \right] \times \Pr\left(V_i + \epsilon_i \geq \max_{j \neq i} \{V_j + \epsilon_j\} \right)
\end{aligned}$$

We proceed in exactly same fashion as above and let $\Delta_{ij} \equiv V_i - V_j$, and $f_\epsilon(\bullet \mid \boldsymbol{\mu}, \sigma)$ be the *joint* PDF of taste shocks $(\epsilon_i, \ldots, \epsilon_n)$ where $\boldsymbol{\mu}$ is the vector of location parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ allowed to differ across ϵ_i, and σ is the *common* scale parameter. The probability of choosing alternative i can then be written as

$$\begin{aligned}
\pi_i &= \Pr\left(V_i + \epsilon_i \geq \max_{j \neq i} \{V_j + \epsilon_j\} \right) \\
&= \Pr\left(\Delta_{ij} + \epsilon_i \geq \max_{j \neq i} \{\epsilon_j\} \right) \qquad (8) \\
&= \Pr\left(\bigcap_{j \neq i} \left\{ \Delta_{ij} + \epsilon_i \geq \epsilon_j \right\} \right) \\
&= \int_{\underline{\epsilon}}^{\overline{\epsilon}} \int_{\underline{\epsilon}}^{\Delta_{i1}+\epsilon_i} \cdots \int_{\underline{\epsilon}}^{\Delta_{in}+\epsilon_i} f_\epsilon(\epsilon_1, \ldots, \epsilon_n \mid \boldsymbol{\mu}, \sigma)\, d\epsilon_1 \cdots d\epsilon_n \\
&= \int_{\underline{\epsilon}}^{\overline{\epsilon}} \int_{\underline{\epsilon}}^{\Delta_{i1}+\epsilon_i} \cdots \int_{\underline{\epsilon}}^{\Delta_{in}+\epsilon_i} f_\epsilon(\epsilon_1 \mid \mu_1, \sigma) \cdots f_\epsilon(\epsilon_n \mid \mu_n, \sigma)\, d\epsilon_1 \cdots d\epsilon_n \\
&= \int_{\underline{\epsilon}}^{\overline{\epsilon}} \left[\prod_{j \neq i} F_\epsilon(\Delta_{ij} + \epsilon_i \mid \mu_j, \sigma) \right] f_\epsilon(\epsilon_i \mid \mu_i, \sigma)\, d\epsilon_i \qquad (9)
\end{aligned}$$

where we again use

$$\int_{\underline{\epsilon}}^{\Delta_{ij}+\epsilon_i} f_\epsilon(\epsilon_j \mid \mu_j, \sigma)\, d\epsilon_j = F_\epsilon(\Delta_{ij} + \epsilon_i \mid \mu_j, \sigma)$$

This is the multivariate analogue of (4). Next, the conditional expectation of alternative i

is

$$
\begin{aligned}
&\mathbf{E}\left[V_i + \epsilon_i \middle| \, V_i + \epsilon_i \geq \max_{j \neq i} \left\{ V_j + \epsilon_j \right\} \right] \\
&\quad = \frac{1}{\pi_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} \int_{\underline{\epsilon}}^{\Delta_{i1}+\epsilon_i} \cdots \int_{\underline{\epsilon}}^{\Delta_{in}+\epsilon_i} \left(V_i + \epsilon_i \right) f_\epsilon \left(\epsilon_1, \ldots, \epsilon_n \,|\, \boldsymbol{\mu}, \sigma \right) d\epsilon_1 \cdots d\epsilon_n \\
&\quad = \frac{1}{\pi_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} \left(V_i + \epsilon_i \right) \left[\prod_{j \neq i} F_\epsilon \left(\Delta_{ij} + \epsilon_i \,|\, \mu_j, \sigma \right) \right] f_\epsilon \left(\epsilon_i \,|\, \mu_i, \sigma \right) d\epsilon_i \\
&\quad = V_i + \frac{1}{\pi_i} \int_{\underline{\epsilon}}^{\overline{\epsilon}} \epsilon_i \left[\prod_{j \neq i} F_\epsilon \left(\Delta_{ij} + \epsilon_i \,|\, \mu_j, \sigma \right) \right] f_\epsilon \left(\epsilon_i \,|\, \mu_i, \sigma \right) d\epsilon_i
\end{aligned}
$$

2.2 Type-I extreme value distribution

As in the binary case, we now turn to the specific case when all ϵ_i follow the Gumbel distribution with CDF (5) and PDF given in (6). To compute expectations, we need to find an expression for

$$
\left[\prod_{j \neq i} F_\epsilon \left(\Delta_{ij} + \epsilon \,|\, \mu_j, \sigma \right) \right] f_\epsilon \left(\epsilon_i \,|\, \mu_i, \sigma \right)
$$

for the Gumbel case:

$$
\begin{aligned}
&\left[\prod_{j \neq i} F_\epsilon \left(\Delta_{ij} + \epsilon \,|\, \mu_j, \sigma \right) \right] f_\epsilon \left(\epsilon_i \,|\, \mu_i, \sigma \right) \\
&\quad = \left[\prod_{j \neq i} \exp \left\{ -e^{-(\Delta_{ij} - \mu_j)/\sigma} \right\} \right] \times \frac{1}{\sigma} \exp \left\{ - \left(\frac{\epsilon_i - \mu_i}{\sigma} \right) - e^{-(\epsilon_i - \mu_i)/\sigma} \right\} \\
&\quad = \frac{1}{\sigma} \exp \left\{ - \sum_{j \neq i} e^{-(\Delta_{ij} + \epsilon_i - \mu_j)/\sigma} \right\} \exp \left\{ - \left(\frac{\epsilon_i - \mu_i}{\sigma} \right) - e^{-(\epsilon_i - \mu_i)/\sigma} \right\} \\
&\quad = \frac{1}{\sigma} \exp \left\{ -e^{-(\epsilon_i - \mu_i)/\sigma} \sum_{j \neq i} e^{-(\Delta_{ij} + (\mu_i - \mu_j))/\sigma} \right\} \exp \left\{ - \left(\frac{\epsilon_i - \mu_i}{\sigma} \right) - e^{-(\epsilon_i - \mu_i)/\sigma} \right\} \\
&\quad = \frac{1}{\sigma} \exp \left\{ - \left(\frac{\epsilon_i - \mu_i}{\sigma} \right) - e^{-(\epsilon_i - \mu_i)/\sigma)} \left[1 + \sum_{j \neq i} e^{-(\Delta_{ij} + \Delta\mu_{ij})/\sigma} \right] \right\}
\end{aligned}
$$

Here we use the definition $\Delta\mu_{ij} \equiv \mu_i - \mu_j$. As in the binary case, we introduce the definition

$$
\alpha_i \equiv \left[1 + \sum_{j \neq i} e^{-(\Delta_{ij} + \Delta\mu_{ij})/\sigma} \right] = \sum_j e^{-(\Delta_{ij} + \Delta\mu_{ij})/\sigma}
$$

such that we can continue simplifying the above expression:

$$\left[\prod_{j\neq i} F_\epsilon\left(\Delta_{ij}+\epsilon \mid \mu_j,\sigma\right)\right] f_\epsilon\left(\epsilon_i \mid \mu_i,\sigma\right)$$
$$= \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i}{\sigma}\right)-e^{-(\epsilon_i-\mu_i)/\sigma)}\alpha_i\right\}$$
$$= \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i-\sigma\log\alpha_i}{\sigma}\right)-\log\alpha_i-e^{-(\epsilon_i-\mu_i)/\sigma)}e^{\log\alpha_i}\right\}$$
$$= \frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\mu_i-\sigma\log\alpha_i}{\sigma}\right)-\log\alpha_i-e^{-(\epsilon_i-\mu_i-\sigma\log\alpha_i)/\sigma)}\right\}$$
$$= \frac{1}{\alpha_i}\times\frac{1}{\sigma}\exp\left\{-\left(\frac{\epsilon_i-\widehat{\mu}_i}{\sigma}\right)-e^{-(\epsilon_i-\widehat{\mu}_i)/\sigma)}\right\}$$
$$= \frac{1}{\alpha_i}\times f_\epsilon\left(\epsilon_i \mid \widehat{\mu}_i,\sigma\right)$$

Here we have defined $\widehat{\mu}_i$ as

$$\widehat{\mu}_i = \mu_i+\sigma\log\alpha_i = \mu_i+\sigma\log\left[\sum_j e^{-(\Delta_{ij}+\Delta\mu_{ij})/\sigma}\right]$$

We consequently get an expression that is identical to the binary case except for the different definition of α_i (and hence $\widehat{\mu}_i$). Thus the probability of choosing alternative i in (9) for the Gumbel case simplifies to

$$\pi_i = \int_{-\infty}^{\infty}\left[\prod_{j\neq i} F_\epsilon\left(\Delta_{ij}+\epsilon \mid \mu_j,\sigma\right)\right] f_\epsilon\left(\epsilon_i \mid \mu_i,\sigma\right) = \frac{1}{\alpha_i}\int_{-\infty}^{\infty} f_\epsilon\left(\epsilon_i \mid \widehat{\mu}_i,\sigma\right)d\epsilon_i = \frac{1}{\alpha_i}$$

and hence

$$\pi_i = \frac{1}{\alpha_i} = \left[\sum_j e^{-(\Delta_{ij}+\Delta\mu_{ij})/\sigma}\right]^{-1} = \left[\sum_j e^{-(V_i-V_j+\mu_i-\mu_j)/\sigma}\right]^{-1} = \frac{e^{(V_i+\mu_i)/\sigma}}{\sum_j e^{(V_j+\mu_j)/\sigma}} \tag{10}$$

This of course is the expression for the probability of choosing alternative i in the multinomial Logit model. The modified location parameter $\widehat{\mu}_i$ can be alternatively written as

$$\widehat{\mu}_i = \mu_i+\sigma\log\alpha_i = \mu_i+\sigma\log\pi_i^{-1} = \mu_i-\sigma\log\pi_i$$

Next, we obtain the following expression for the expected value of ϵ_i conditional on i being the alternative with the highest value:

$$\int_{\underline{\epsilon}}^{\overline{\epsilon}}\epsilon_i\left[\prod_{j\neq i} F_\epsilon\left(\Delta_{ij}+\epsilon \mid \mu_j,\sigma\right)\right] f_\epsilon\left(\epsilon_i \mid \mu_i,\sigma\right) = \frac{1}{\alpha_i}\int_{-\infty}^{\infty}\epsilon_i f_\epsilon\left(\epsilon_i \mid \widehat{\mu}_i,\sigma\right)d\epsilon_i$$
$$= \pi_i\left[\widehat{\mu}_i+\sigma\gamma\right] = \pi_i\left[\mu_i-\sigma\log\pi_i+\sigma\gamma\right]$$

which is identical to the binary-choice case in (7). Using these results, the expected value function is given by

$$\mathbf{E}V = \sum_i \pi_i \left[V_i + \mu_i - \sigma \log \pi_i + \sigma\gamma \right]$$

Since,

$$\sigma \log \pi_i = \sigma \log \left(e^{(V_i+\mu_i)/\sigma} \right) - \sigma \log \left(\sum_j e^{(V_j+\mu_j)/\sigma} \right) = V_i + \mu_i - \sigma \log \left(\sum_j e^{(V_j+\mu_j)/\sigma} \right)$$

we can further simply the expectation to obtain

$$\mathbf{E}V = \sum_i \pi_i \left[\sigma \log \left(\sum_j e^{(V_j+\mu_j)/\sigma} \right) + \sigma\gamma \right] = \sigma \log \left(\sum_j e^{(V_j+\mu_j)/\sigma} \right) + \sigma\gamma \tag{11}$$

Finally, if we additionally impose $\mu_i = \mu$ for all j we get the more standard expression

$$\begin{aligned} \mathbf{E}V = \sigma \log \left(\sum_j e^{(V_j+\mu)/\sigma} \right) + \sigma\gamma &= \sigma \log \left(e^{\mu/\sigma} \sum_j e^{V_j/\sigma} \right) + \sigma\gamma \\ &= \sigma \log \left(\sum_j e^{V_j/\sigma} \right) + \mu + \sigma\gamma \end{aligned}$$

A Derivatives

A.1 Multinomial choice model

In numerical applications, we often need the first (and possibly second) derivative of $\mathbf{E}V(\boldsymbol{x})$ w.r.t. some element of $\boldsymbol{x}$. We derive these expressions in this section, assuming throughout that the taste shocks are drawn from a Gumbel distribution with potentially different location parameters. The first-order derivative w.r.t. some x_k is given by

$$\frac{\partial \mathbf{E}V(\boldsymbol{x})}{\partial x_k} = \sigma \sum_i \left[\frac{e^{(V_i+\mu_i)/\sigma}}{\sum_j e^{(V_j+\mu_j)/\sigma}} \frac{1}{\sigma} \frac{\partial V_i(\boldsymbol{x})}{\partial x_k} \right] = \sum_i \pi_i \frac{\partial V_i(\boldsymbol{x})}{\partial x_k}$$

which is thus just a probability-weighted average of all V_i derivatives. Next, the second-order derivative involves derivatives w.r.t. the probabilities π_i, which we obtain first:

$$\begin{aligned}
\frac{\partial \pi_i(\boldsymbol{x})}{\partial x_k} &= \frac{\partial}{\partial x_k}\left[\frac{e^{(V_i(\boldsymbol{x})+\mu_i)/\sigma}}{\sum_j e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}}\right] \\
&= \frac{1}{\sigma}\frac{e^{(V_i+\mu_i)/\sigma}}{\sum_j e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}}\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \frac{1}{\sigma}\frac{e^{(V_i+\mu_i)/\sigma}}{\left(\sum_j e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}\right)^2}\sum_j\left[e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}\frac{\partial V_j(\boldsymbol{x})}{\partial x_k}\right] \\
&= \frac{1}{\sigma}\frac{e^{(V_i+\mu_i)/\sigma}}{\sum_j e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}}\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \frac{1}{\sigma}\frac{e^{(V_i+\mu_i)/\sigma}}{\sum_j e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}}\sum_j\left[\frac{e^{(V_j(\boldsymbol{x})+\mu_j)/\sigma}}{\sum_m e^{(V_m(\boldsymbol{x})+\mu_m)/\sigma}}\frac{\partial V_j(\boldsymbol{x})}{\partial x_k}\right] \\
&= \frac{1}{\sigma}\pi_i\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \frac{1}{\sigma}\pi_i\sum_j \pi_j\frac{\partial V_j(\boldsymbol{x})}{\partial x_k} \\
&= \frac{1}{\sigma}\pi_i\left[\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \sum_j \pi_j\frac{\partial V_j(\boldsymbol{x})}{\partial x_k}\right]
\end{aligned}$$

The second-order derivative of $\mathbf{E}V(\boldsymbol{x})$ is then given by

$$\frac{\partial^2 \mathbf{E}V(\boldsymbol{x})}{\partial x_k^2} = \sum_i\left[\frac{\partial \pi_i(\boldsymbol{x})}{\partial x_k}\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} + \pi_i(\boldsymbol{x})\frac{\partial^2 V_i(\boldsymbol{x})}{\partial x_k^2}\right]$$

Summing-up restriction. Note that since the probabilities of choosing alternatives must sum to unity, their derivatives have to sum to zero. We can verify that this restriction holds for the above expression:

$$\begin{aligned}
\sum_i\frac{\partial \pi_i(\boldsymbol{x})}{\partial x_k} &= \frac{1}{\sigma}\sum_i \pi_i\left[\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \sum_j \pi_j\frac{\partial V_j(\boldsymbol{x})}{\partial x_k}\right] \\
&= \frac{1}{\sigma}\left[\sum_i \pi_i\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \sum_i \pi_i\sum_j \pi_j\frac{\partial V_j(\boldsymbol{x})}{\partial x_k}\right] \\
&= \frac{1}{\sigma}\left[\sum_i \pi_i\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \left(\sum_i \pi_i\right)\left(\sum_j \pi_j\frac{\partial V_j(\boldsymbol{x})}{\partial x_k}\right)\right] = 0
\end{aligned}$$

A.2 Binary choice model

In the binary case, the above expressions simplify to the following:

$$\begin{aligned}
\frac{\partial \mathbf{E}V(\boldsymbol{x})}{\partial x_k} &= \pi_1\frac{\partial V_1(\boldsymbol{x})}{\partial x_k} + (1-\pi_1)\frac{\partial V_2(\boldsymbol{x})}{\partial x_k} \\
\frac{\partial \pi_i(\boldsymbol{x})}{\partial x_k} &= \frac{1}{\sigma}\pi_i(1-\pi_i)\left[\frac{\partial V_i(\boldsymbol{x})}{\partial x_k} - \frac{\partial V_{-i}(\boldsymbol{x})}{\partial x_k}\right] = -\frac{\partial \pi_{-i}(\boldsymbol{x})}{\partial x_k}
\end{aligned}$$

Futhermore, for the second-order derivative we find that

$$\begin{aligned}
\frac{\partial^2 \mathbf{E}V(\boldsymbol{x})}{\partial x_k^2} &= \left[\frac{\partial \pi_1(\boldsymbol{x})}{\partial x_k}\frac{\partial V_1(\boldsymbol{x})}{\partial x_k} + \pi_1 \frac{\partial^2 V_1(\boldsymbol{x})}{\partial x_k^2}\right] + \left[\frac{\partial \pi_2(\boldsymbol{x})}{\partial x_k}\frac{\partial V_2(\boldsymbol{x})}{\partial x_k} + \pi_2 \frac{\partial^2 V_2(\boldsymbol{x})}{\partial x_k^2}\right] \\
&= \left[\frac{\partial \pi_1(\boldsymbol{x})}{\partial x_k}\frac{\partial V_1(\boldsymbol{x})}{\partial x_k} + \pi_1 \frac{\partial^2 V_1(\boldsymbol{x})}{\partial x_k^2}\right] + \left[-\frac{\partial \pi_1(\boldsymbol{x})}{\partial x_k}\frac{\partial V_2(\boldsymbol{x})}{\partial x_k} + \pi_2 \frac{\partial^2 V_2(\boldsymbol{x})}{\partial x_k^2}\right] \\
&= \frac{\partial \pi_1(\boldsymbol{x})}{\partial x_k}\left[\frac{\partial V_1(\boldsymbol{x})}{\partial x_k} - \frac{\partial V_2(\boldsymbol{x})}{\partial x_k}\right] + \pi_1 \frac{\partial^2 V_1(\boldsymbol{x})}{\partial x_k^2} + (1-\pi_1)\frac{\partial^2 V_2(\boldsymbol{x})}{\partial x_k^2} \\
&= \frac{\sigma}{\pi_1(1-\pi_1)}\left(\frac{\partial \pi_1(\boldsymbol{x})}{\partial x_k}\right)^2 + \pi_1 \frac{\partial^2 V_1(\boldsymbol{x})}{\partial x_k^2} + (1-\pi_1)\frac{\partial^2 V_2(\boldsymbol{x})}{\partial x_k^2}
\end{aligned}$$

B Numerical implementation

The expressions derived in the previous sections are not necessarily suited to be used in numerical implementations as is, since they can yield undefined results (i.e. NaNs). For example, the probability of choosing alternative i in (10) can result in an expression of the form "∞/∞" in numerical applications if V_i is large, as the exponential function quickly approaches values too large to be represented as floating-point numbers. One alternative way to compute the probabilities in a numerically stable way is to define

$$m = \arg\max_j \left\{V_j + \mu_j\right\}$$

and then divide both numerator and denominator by $e^{(V_m+\mu)/\sigma}$,

$$\pi_i = \frac{e^{(V_i - V_m + \mu_i - \mu_m)/\sigma}}{\sum_j e^{(V_j - V_m + \mu_j - \mu_m)/\sigma}} = \frac{e^{(V_i - V_m + \mu_i - \mu_m)/\sigma}}{1 + \sum_{j\neq m} e^{(V_j - V_m + \mu_j - \mu_m)/\sigma}}$$

This way the numerator is guaranteed to be in $[0, 1]$, and the denominator will be in $[1, n]$. Furthermore, as long as V_m is finite, there won't be any arithmetic operations of the "$\infty - \infty$" sort, while an infinite V_m means that all choices yield $-\infty$ utility, which should not arise.

Similarly, the expression for $\mathbf{E}V$ in (11)

$$\mathbf{E}V = \sigma \log\left(\sum_j e^{(V_j+\mu_j)/\sigma}\right) + \sigma\gamma$$

returns ∞ in standard implementations even for moderately large V_i, even though the true value is finite. In fact, it is easy to see that the upper bound is given by

$$\begin{aligned}
\mathbf{E}V &= \sigma \log\left(\sum_j e^{(V_j+\mu_j)/\sigma}\right) + \sigma\gamma \\
&\leq \sigma \log\left(\sum_j e^{(V_m+\mu_m)/\sigma}\right) + \sigma\gamma \\
&= \sigma \log\left(n \cdot e^{(V_m+\mu_m)/\sigma}\right) + \sigma\gamma \\
&= \sigma \log n + V_m + \mu_m + \sigma\gamma
\end{aligned}$$

This suggests the following alternative expression:

$$\begin{aligned}
\mathbf{E}V &= \sigma \log\left(e^{(V_m+\mu_m)/\sigma} \sum_j e^{(V_j-V_m+\mu_j-\mu_m)/\sigma}\right) + \sigma\gamma \\
&= V_m + \mu_m + \sigma \log\left(\sum_j e^{(V_j-V_m+\mu_j-\mu_m)/\sigma}\right) + \sigma\gamma \\
&= V_m + \mu_m + \sigma \log\left(1 + \sum_{j\neq m} e^{(V_j-V_m+\mu_j-\mu_m)/\sigma}\right) + \sigma\gamma
\end{aligned}$$

In particular, for the binary model this reduces to

$$\mathbf{E}V = V_m + \mu_m + \sigma \log\left(1 + e^{(V_{-m}-V_m+\mu_{-m}-\mu_m)/\sigma}\right) + \sigma\gamma$$

Additionally, since all the terms in the sum

$$z \equiv \sum_{j\neq m} e^{(V_j-V_m+\mu_j-\mu_m)/\sigma}$$

are small by construction, one can use the `log1p` function to evaluate $\log(1+z)$ numerically.