Some Useful Extreme Value Derivations

Chase Abram*

December 3, 2023

1 Purpose

To model choices over a set of options, it is often useful to say that agents receive idiosyncratic preference shocks over the choices. This assumption delivers realistic "smoothness" in outcomes, in that choices with better fundamentals are not chosen by every agent, since preference shocks cause many (often most) agents to select an option with submaximal fundamentals (but of course larger personal preference). The two dominant ways of modelling this assumption are via assuming an additive preference shock or a multiplicative preference shock. For an additive shock, the convenient shock distribution is Gumbel, and for the multiplicative shock, the convenient shock distribution is Frechet.

All the results below are canonical, but I often find myself getting confused (e.g. should this be a $\exp(\epsilon A)$ or A^ϵ form?), so I decided to collect all the algebra and results in one place for a handy reference. I first show some properties of using Gumbel shocks, then show the relationship between Frechet and Gumbel shocks so I can also show properties of using Frechet shocks. I briefly explain how the finite choice case extends to infinitely many choices, then provide some examples and nonexamples.

Tildes denote logs, i.e. $\tilde{A} \equiv \ln A$. It may then seem strange to start below by defining tilde variables, but it will become apparent that this is the correct notion, since logs of Frechet random variables are Gumbel.

*Send corrections or comments to abram@uchicago.edu.

2 Summary

For quick reference, I list the main results from below.

- If $\tilde{\eta} \sim \text{Gumbel}(\mu, \beta)$, then $\mathbb{E}[\eta] = \mu + \beta\gamma$, where γ is the Euler-Mascheroni constant
- If $\tilde{\eta} \sim \text{Gumbel}(0, \frac{1}{\epsilon})$, $\tilde{A}$ is real, then $\tilde{A} + \tilde{\eta} \sim \text{Gumbel}(\tilde{A}, \frac{1}{\epsilon})$
- If $\tilde{\eta}_i \sim \text{Gumbel}(0, \frac{1}{\epsilon})$ iid, $\tilde{A}_i$ are real, then $\max_i\{\tilde{A}_i + \tilde{\eta}_i\} \sim \text{Gumbel}\left(\ln\left(\sum_{i=1}^n \exp(\epsilon\tilde{A}_i)\right), \frac{1}{\epsilon}\right)$
- If $\tilde{\eta}_i \sim \text{Gumbel}(0, \frac{1}{\epsilon})$ iid, $\tilde{A}_i$ are real, then $Pr[j = \text{argmax}_i\{\tilde{A}_i + \tilde{\eta}_i\} = \frac{\exp(\epsilon\tilde{A}_j)}{\sum_i \exp(\epsilon\tilde{A}_i)}$
- If $\eta \sim \text{Frechet}(\epsilon)$, A is positive, then $\ln(A\eta) \sim \text{Gumbel}(\ln A, \frac{1}{\epsilon})$
- If $\eta_i \sim \text{Frechet}(\epsilon)$ iid, A_i are positive, then $\max_i\{\ln(A_i\eta_i)\} \sim \text{Gumbel}(\ln\sum_i A_i^\epsilon, \frac{1}{\epsilon})$
- If $\eta_i \sim \text{Frechet}(\epsilon)$ iid, A_i are positive, then $Pr[j = \text{argmax}_i\, A_i\eta_i] = \frac{A_j^\epsilon}{\sum_i A_i^\epsilon}$

3 Gumbel

I start with Gumbel shocks (the additive form). The ultimate result is to find the choice probabilities, but to get there I first show some intermediate results. The final result requires a beast of an integral, but luckily some eagle-eyed substitution makes the problem quite easy, and the final answer intuitive.

3.1 Definition

The Gumbel(μ, β) distribution with location μ and scale β has the following CDF F and PDF f.

$$\begin{aligned}
F(x) &= \exp\left(-\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\right) \\
f(x) \equiv F'(x) &= \frac{1}{\beta}\exp\left(-\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\right)\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)
\end{aligned}$$

The mean of such a distribution may be calculated by employing a variety of simple, but perhaps unintuitive to guess, calculus steps. First, use the substitution $z = \frac{x-\mu}{\beta}$, then use the substitution $y = e^{-z}$. These two substitutions simplify the integral significantly, and allow for evaluating in two pieces. The second piece is a trivial improper integral. The first piece's integrand may be recognized as the derivative of a particular function evaluated at zero, then the differentiation operator may be pulled outside of the integral. The remaining integral is simply a Gamma function, and using the definition of the Euler-Mascheroni constant $\gamma \equiv \Gamma'(1)$ yields the final solution. I complete these steps below.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x\frac{1}{\beta}\exp\left(-\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\right)\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\mathrm{d}x && \text{(Plug in)} \\
&= \int_{-\infty}^{\infty} (\beta z + \mu)\exp\left(-\exp\left(-\left(z\right)\right)\right)\exp\left(-\left(z\right)\right)\mathrm{d}z && (z = \tfrac{x-\mu}{\beta}) \\
&= \int_{0}^{\infty} (\beta \ln y + \mu)\exp\left(-y\right)\mathrm{d}y && (y = e^{-z}) \\
&= \int_{0}^{\infty} \beta \ln y \exp\left(-y\right)\mathrm{d}y + \int_{0}^{\infty} \mu \exp\left(-y\right)\mathrm{d}y && \text{(Split up pieces)} \\
&= \beta\int_{0}^{\infty} \left[\frac{\mathrm{d}}{\mathrm{d}k} y^k \exp\left(-y\right)\Big|_{k=0}\right]\mathrm{d}y + \mu\left[-e^{-y}\Big|_0^{\infty}\right] && \text{(Calc rules)} \\
&= \beta\frac{\mathrm{d}}{\mathrm{d}k}\Gamma(k+1)\Big|_{k=0} + \mu && \text{(Def. of } \Gamma) \\
&= \beta\gamma + \mu
\end{aligned}$$

3.2 Additive shift

Let $X \equiv \tilde{\eta} \sim \text{Gumbel}(0, \frac{1}{\epsilon})$, so it has the following distribution.

$$F_X(x) = \exp\left(-\exp\left(-\epsilon x\right)\right)$$

Let $\tilde{A}$ be a real number. Then $Y \equiv \tilde{A} + \tilde{\eta}$ is Gumbel$(\tilde{A}, \frac{1}{\epsilon})$:

$$
\begin{aligned}
F_Y(x) &= Pr[\tilde{A} + \tilde{\eta} \leq x] \\
&= Pr[\tilde{\eta} \leq x - \tilde{A}] \\
&= \exp\left(-\exp\left(-\epsilon(x - \tilde{A})\right)\right)
\end{aligned}
$$

3.3 Maximum of n additively shifted independent draws

Now consider $Y_i \equiv \tilde{A}_i + \tilde{\eta}_i \sim$ Gumbel$(\tilde{A}_i, \frac{1}{\epsilon})$ for n real numbers $\tilde{A}_i$ and n independent $\tilde{\eta}_i \sim$ Gumbel$(0, \frac{1}{\epsilon})$. The CDF of $\max_i\{\tilde{A}_i + \tilde{\eta}_i\}$ is:

$$
\begin{aligned}
F_{\max_i Y_i}(x) &= Pr[\max_i\{\tilde{A}_i + \tilde{\eta}_i\} \leq x] \\
&= \prod_{i=1}^{n} Pr[\tilde{A}_i + \tilde{\eta}_i \leq x] && \text{(Independent)} \\
&= \prod_{i=1}^{n} \exp\left(-\exp\left(-\epsilon(x - \tilde{A}_i)\right)\right) \\
&= \exp\left(-\sum_{i=1}^{n} \exp\left(-\epsilon(x - \tilde{A}_i)\right)\right)
\end{aligned}
$$

This is decent-looking, but it's unclear if this is a Gumbel distribution or not. The trick is to recognize that we can log the sum and move it into the inner exponential, and see $\max_i\{\tilde{A}_i + \eta_i\} \sim$ Gumbel $\left(\ln\left(\sum_{i=1}^{n}\exp(\epsilon\tilde{A}_i)\right), \frac{1}{\epsilon}\right)$:

$$
\begin{aligned}
\exp\left(-\sum_{i=1}^{n} \exp\left(-\epsilon(x - \tilde{A}_i)\right)\right) &= \exp\left(-\exp(-\epsilon x)\sum_{i=1}^{n} \exp\left(\epsilon\tilde{A}_i\right)\right) \\
&= \exp\left(-\exp\left(-\epsilon(x - \ln\left(\sum_{i=1}^{n}\exp(\epsilon\tilde{A}_i)\right)\right)\right)
\end{aligned}
$$

3.4 Selection probabilities for max of n independent shifted draws

Now suppose we are interested in knowing the probability that $\tilde{A}_j + \tilde{\eta}_j$ is the maximum of the n choices, in other words $j = \text{argmax}_i\{\tilde{A}_i + \tilde{\eta}_i\}$. Call this probability π_j.

$$
\begin{aligned}
\pi_j &= Pr[\tilde{A}_j + \tilde{\eta}_j = \max_i\{\tilde{A}_i + \tilde{\eta}_i\}] \\
&= Pr[\max_{i \neq j}\{\tilde{A}_i + \tilde{\eta}_i\} \leq \tilde{A}_j + \tilde{\eta}_j] \\
&= \int_{\mathbb{R}} Pr[\max_{i \neq j}\{\tilde{A}_i + \tilde{\eta}_i\} \leq z] Pr[\tilde{A}_j + \tilde{\eta}_j = z] \mathrm{d}z
\end{aligned}
$$

The last line above is perhaps a bit fast and loose with the measure theory, but nonetheless it makes clear what we mean: we consider the probability that all other draws are less than some value which

j takes, then sum over the probability that j takes each possible value. The above results are now useful to replace the Pr terms in the integrand, where the second term is the density of $\tilde{A}_j + \tilde{\eta}_j$

$$\pi_j = \int_{\mathbb{R}} \exp\left(-\exp\left(-\epsilon(z - \ln\left(\sum_{i \neq j} \exp(\epsilon \tilde{A}_i)\right)\right)\right) \epsilon \exp\left(-(\epsilon(z - \tilde{A}_j) + \exp\left(-\epsilon(z - \tilde{A}_j)\right)\right) \mathrm{d}z$$

What a mess! Let's move some things around to try to clean it up.

$$\pi_j = \int_{\mathbb{R}} \exp\left(-\exp\left(-\epsilon\left(z - \ln \sum_i \exp(\epsilon \tilde{A}_i)\right)\right)\right) \epsilon \exp\left(-\epsilon(z - \tilde{A}_j)\right) \mathrm{d}z$$

Closer, but the integral still seems so ugly. This last step is the least intuitive, but what we can do is recognize that the integrand looks similar to the density of a Gumbel. In particular, the integral is in the following form, where $y \equiv z - \tilde{A}_j$.

$$\int_{\mathbb{R}} e^{-e^{-\epsilon y} \cdot \frac{\sum_i \exp(\epsilon \tilde{A}_i)}{\exp(\epsilon \tilde{A}_j)}} \epsilon e^{-\epsilon y} \mathrm{d}y$$

Letting $u = e^{-e^{-\epsilon y} \cdot \frac{\sum_i \exp(\epsilon \tilde{A}_i)}{\exp(\epsilon \tilde{A}_j)}}$, we have $\mathrm{d}u = \frac{\sum_i \exp(\epsilon \tilde{A}_i)}{\exp(\epsilon \tilde{A}_j)} e^{-e^{-\epsilon y} \cdot \frac{\sum_i \exp(\epsilon \tilde{A}_i)}{\exp(\epsilon \tilde{A}_j)}} \epsilon e^{-\epsilon y} \mathrm{d}y$, and thus the integral is simply

$$\begin{aligned}\pi_j &= \int_0^1 \frac{\exp(\epsilon \tilde{A}_j)}{\sum_i \exp(\epsilon \tilde{A}_i)} \mathrm{d}u \\ &= \frac{\exp(\epsilon \tilde{A}_j)}{\sum_i \exp(\epsilon \tilde{A}_i)}\end{aligned}$$

4 Frechet

Rather than rederive all the corresponding above results for the Frechet shock case, I just show how to transform the Frechet form into the Gumbel form, then get the results directly.

4.1 Definition

The Frechet(ϵ) distribution with shape ϵ has the following CDF.

$$F(x) = \exp\left(-x^{-\epsilon}\right)$$

Let $X \sim$ Frechet(ϵ), and define $Y \equiv \ln X$. Then $Y \sim$ Gumbel($0, \frac{1}{\epsilon}$):

$$\begin{aligned}
F_Y(y) &= Pr[Y \leq y] \\
&= Pr[\ln X \leq y] \\
&= Pr[X \leq \exp(y)] \\
&= F(\exp(y)) \\
&= \exp\left(-\left(\exp(y)\right)^{-\epsilon}\right) \\
&= \exp\left(-\exp\left(-\epsilon y\right)\right)
\end{aligned}$$

Therefore we can log the Frechet variables of interest to get them in Gumbel form, then apply the above Gumbel results, without having to redo all the math.

4.2 Multiplicative shift

Let $\eta \sim$ Frechet(ϵ). Then $\ln(A\eta) = \tilde{A} + \tilde{\eta}$, thus $\ln(A\eta) \sim$ Gumbel($\tilde{A}, \frac{1}{\epsilon}$).

4.3 Maximum of n multiplicatively shifted independent draws

Let A_i be n positive real numbers and η_i be n independent Frechet(ϵ) draws. Then $\max_i\{\ln(A_i\eta_i)\} = \max_i\{\tilde{A}_i + \tilde{\eta}_i\} \sim$ Gumbel($\ln \sum_{i=1}^n \exp(\epsilon \tilde{A}_i), \frac{1}{\epsilon}$) $=$ Gumbel($\ln \sum_{i=1}^n A_i^\epsilon, \frac{1}{\epsilon}$).

4.4 Selection probabilities for max of n indepdenent multiplicatively shifted draws

Monotonic transformations preserve ordering, so $Pr[A_j\eta_j = \max_i\{A_i\eta_i\}] = Pr[\tilde{A}_j + \tilde{\eta}_j = \max_i\{\tilde{A}_i + \tilde{\eta}_i\}$. Therefore, if we let π_j now denote the probability $A_j\eta_j$ is the max, we have

$$\begin{aligned}
\pi_j &= \frac{\exp(\epsilon \tilde{A}_j)}{\sum_{i=1}^n \exp(\epsilon \tilde{A}_i)} \\
&= \frac{A_j^\epsilon}{\sum_{i=1}^n A_i^\epsilon}
\end{aligned}$$

5 Infinitely Many Choices

The above exposition assumed finitely many locations, and thus sidestepped any concerns about sums converging. Nonetheless, the results above go through as long as the relevant sums or integrals converge. For countably many choices indexed by $\aleph$, the choice probabilities are simply

$$\pi_j = \frac{\exp(\epsilon \tilde{A}_j)}{\sum_{i \in \aleph} \exp(\epsilon \tilde{A}_j)}$$

The same idea holds for a continuum of choices C, we just replace the sums with integrals.

$$\pi(j) = \frac{\exp(\epsilon \tilde{A}(j))}{\int_C \exp(\epsilon \tilde{A}(i)) \mathrm{d}i}$$

The slight complication is that the $\pi(j)$ are now with respect to whatever measure C is endowed with, whereas above we were implicitly always using the counting measure. For more on the exact technical details of the construction of the shock process to make this representation make sense, see Cosslett (1988).

6 Clarifying Examples

6.1 Finite choices spatial example with Gumbel

Consider a unit measure of households that all live at the same location, and are each choosing where to work, among n locations. Location i takes $\tilde{\tau}_i$ minutes to commute to, but each household draws an independent preference draw for each location, with dispersion parameter $\frac{1}{\epsilon}$, so their potential utilities are $-\tilde{\tau}_i + \tilde{\eta}_i$, and each household picks the location which delivers maximal utility. Then the share of households choosing j is

$$\pi_j = \frac{\exp(-\epsilon\tilde{\tau}_j)}{\sum_{i=1}^n \exp(-\epsilon\tilde{\tau}_i)}$$

Locations which take longer to commute to are chosen by a lower share of households. Additionally, as ϵ increases, shocks are less dispersed, so a higher share of households go to the lowest $\tilde{\tau}$ location.

6.2 Finite choice spatial example with Frechet

Now suppose households get utility from each location, but that if the location is τ away, they only get $\frac{1}{\tau}\eta$ utility, where η is iid Frechet(ϵ) across all locations and households. Then the share of households choosing j is

$$\begin{aligned}\pi_j &= \frac{\exp(-\epsilon\tilde{\tau}_j)}{\sum_{i=1}^n \exp(-\epsilon\tilde{\tau}_i)} \\ &= \frac{\tau_j^{-\epsilon}}{\sum_{i=1}^n \tau_i^{-\epsilon}}\end{aligned}$$

6.3 Countable choices example

Let $i \in \{0, 1, \ldots\}$, and $A_i = \beta^i$, where $\beta \in (0, 1)$. Utility is $A_i\eta_i$, where η_i is iid Frechet(ϵ). Then

$$\begin{aligned}\pi_j &= \frac{A_j^\epsilon}{\sum_i A_i^\epsilon} \\ &= \frac{\beta^{\epsilon j}}{\frac{1}{(1-\beta)^\epsilon}} \\ &= \left((1-\beta)\beta^j\right)^\epsilon\end{aligned}$$

The A_i are decaying fast enough that the sum in the denominator converges, so the π_j are proper probabilities.

6.4 Countable choices nonexample

Suppose instead $A_i = 1$, again for $i \in \{0, 1, \ldots\}$. Then the denominator will not converge, and thus the π terms are nonsense. The economic reason is that there is no maximal choice, because for any potential optimal choice j, with probability 1 there is a better option $j' > j$. Mathematically, the reason is that the expected value of the max is increasing in n when all the A_i are the same, thus having countably many possibilities will mean the expectation of the max is infinite. This was not an issue in the previous example because, although there were infinitely many choices, the exponential decay of A as j increased meant that high j options offered vanishingly little chance of being maximal, and thus the expectation remained finite.

6.5 Continuum of choices example

Suppose $A(i) = i^{\beta}$ for $i \in [0, 1]$, $\beta > 0$. Then the choice probabilities are

$$\pi(j) = \frac{j^{\epsilon\beta}}{\int_0^1 i^{\epsilon\beta} \mathrm{d}i}$$

where the Lebesgue measure is implicitly being used.

6.6 Continuum of choices nonexample

Suppose $A(i) = \frac{1}{i}$ for $i \in (0, 1]$ and $A(0) = 1$. Then for $j > 0$, the implied choice probabilities would be

$$\pi(j) = \frac{j^{-\epsilon}}{\int_0^1 i^{-\epsilon} \mathrm{d}i}$$

but the integral in the denominator does not converge. The issue is similar to the above counterexample: for any j which might be optimal, with probability 1 there exists a $j' < j$ which delivers a higher $A(j)\eta(j)$. Mathematically, the expectation of the maximum over the continuum simply does not exist.