# Common Distributions

Paolo Zacchia

Statistics and Econometrics

Lecture 2

# General overview

This lecture analyzes the most common univariate distributions that are encountered in statistical analysis and are referenced in this course.

The analysis proceeds as follows.

1. Discrete distributions.

2. Continuous distributions: location-scale families.

3. Continuous distributions: other common families.

4. Continuous distributions: generalized extreme value.

# Parametric families

- The expressions ‘distribution’ and ‘parametric distribution family’ are often conflated.

- A “family” of distributions is a set of distributions that are identical up some numeric *parameters*.

- Example: all experiments about “tossing a coin” belong to the **Bernoulli** family, but the probability $p$ of observing a ‘Head’ might differ between coins with different “balance.”

- Thus $p$ is a **parameter** of a Bernoulli distribution and:

$$X \sim \mathrm{Be}\,(p)$$

  here means: the random variable $X$ follows the Bernoulli (Be) distribution with parameter $p$.

# Overviewing a distribution (1/2)

The analysis of distributions is systematic. For each is reported:

- the support, e.g. $\mathbb{X} \in \{0, 1\}$;
- the parameters and their admissible values, e.g. $p \in [0, 1]$;
- the notation by which they are indicated, e.g. $X \sim \text{Be}(p)$;
- the p.m.f. or p.d.f., e.g. $f_X(x; p) = p^x (1 - p)^{1-x}$;
- the c.d.f., e.g.

$$F_X(x; p) = (1 - p)\mathbb{1}[x \in [0, 1)] + \mathbb{1}[x \in [1, \infty)];$$

- the m.g.f., e.g. $M_X(t; p) = p \exp(t) + (1 - p)$;
- the key moments (usually only mean and variance).

These examples refer to to the Bernoulli distribution (family).

# Overviewing a distribution (2/2)

In addition, it often useful to report:

- some key derivations (of the m.g.f., of key moments, etc.);
- the **graph** of **density functions** (continuous distributions only);
- the **relationships** between a certain distribution and other distributions which result from specific **transformations** or from equating parameters across distributions.

Relationships between distributions are especially useful to link each distribution to the **real world** phenomena that it is meant to describe, and to the underlying **intuition**.

# The Bernoulli distribution

- The Bernoulli distribution describes all dichotomous events.

- The Bernoulli distribution is elementary, and it forms the basis for other discrete distributions.

- Note: with $p = 0$ or $p = 1$, the entire probability mass is on one realization ("degenerate distribution").

- The key moments are:

$$\mathbb{E}[X] = p$$
$$\mathbb{V}\mathrm{ar}[X] = p(1-p)$$

  where $\mathbb{V}\mathrm{ar}[X]$ is maximized at $p = 0.5$ ("balanced" case).

# The binomial distribution (1/3)

- The binomial distribution corresponds to the repetition of $n \in \mathbb{N}$ Bernoulli "experiments" (also called **trials**).

- A random variable $X$ that follows the binomial distribution counts the number of $0 \leq x \leq n$ **successes** of the Bernoulli experiment that defines it (every Bernoulli $x = 1$ counts).

- The underlying sample space is the set $\mathbb{S} = \{0, 1\}^n$. All the underlying Bernoulli trials are independent.

- The support of $X$ is $\mathbb{X} = \{0, 1, \ldots, n\}$.

- The binomial distribution has two parameters, $p$ and $n$.

$$X \sim \mathrm{Bn}\,(p, n)$$

# The binomial distribution (2/3)

- The binomial distribution owes its name to its extensive use of the binomial coefficient and formula. In fact:

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

  are the ways to obtain $x$ successes out of $n$ trials.

- The p.m.f. is thus:

$$f_X(x;p,n) = \binom{n}{x} p^x (1-p)^{n-x}$$

- ...and the c.d.f. is as follows (it equals 1 for $x = n$).

$$F_X(x;p,n) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

# The binomial distribution (3/3)

- By the binomial formula, the distribution's m.g.f. is:

$$\begin{aligned} M_X(t;p,n) &= \sum_{x=0}^{n} \binom{n}{x} \exp(tx)\, p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} [\exp(t)\, p]^x (1-p)^{n-x} \\ &= [p\exp(t) + (1-p)]^n \end{aligned}$$

- ...hence, the mean and variance are:

$$\begin{aligned} \mathbb{E}[X] &= np \\ \mathbb{V}\mathrm{ar}[X] &= np(1-p) \end{aligned}$$

which is intuitive.

# The geometric distribution (1/3)

- The geometric distribution is also based on some **possibly infinite** Bernoulli trials. Implicitly, trials are ordered.

- Here $X$ indexes the trial that delivers the **first success**.

- The support is $\mathbb{X} = \mathbb{N}$ and there is one parameter: $p$.

- Since $x - 1$ failures must occur before a success occurs at $x$, the p.m.f. (that gives the distribution its name) is:

$$f_X(x; p) = p(1-p)^{x-1}$$

- ...and the c.d.f. is as follows (it tends to 1 as $x \to \infty$).

$$F_X(x; p) = \sum_{i=0}^{\lfloor x \rfloor - 1} p(1-p)^i = 1 - (1-p)^{\lfloor x \rfloor}$$

# The geometric distribution (2/3)

- The geometric m.g.f. exists for $t < -\log(1-p)$:

$$\begin{aligned} M_X(t;p) &= \lim_{M\to\infty} \sum_{x=0}^{M} \exp(tx) \cdot p(1-p)^{x-1} \\ &= p\exp(t) \cdot \lim_{M\to\infty} \sum_{x=0}^{M} \left[(1-p)\cdot\exp(t)\right]^{x-1} \\ &= p\exp(t) \cdot \lim_{M\to\infty} \frac{1-\left[(1-p)\cdot\exp(t)\right]^{M}}{1-(1-p)\cdot\exp(t)} \\ &= \frac{p\exp(t)}{1-(1-p)\exp(t)} \end{aligned}$$

- ... hence, the mean and variance are as follows.

$$\mathbb{E}[X] = \frac{1}{p}$$
$$\mathbb{V}\mathrm{ar}[X] = \frac{1-p}{p^2}$$

# The geometric distribution (3/3)

- The geometric distribution has the *memoryless* property.

$$\begin{aligned}\mathbb{P}\left(X > s \mid X > t\right) &= \frac{\mathbb{P}\left(X > s \cap X > t\right)}{\mathbb{P}\left(X > t\right)}\\ &= \frac{\mathbb{P}\left(X > s\right)}{\mathbb{P}\left(X > t\right)}\\ &= (1-p)^{s-t}\\ &= \mathbb{P}\left(X > s-t\right)\end{aligned}$$

- The probability of a success at the $s$-th trial conditional on $t$ failures is equal to the **ex ante** probability that a success occurs after exactly $s - t$ trials.

- Every failure "resets" the probability of a success!

# The negative binomial distribution (1/2)

- The **negative** binomial models the occurrence of the first $r \in \mathbb{N}$ successes out of a series of Bernoulli trials.

- Here $X$ indexes the trial that delivers the $\boldsymbol{r}$**-th success**.

- The support is $\mathbb{X} = \mathbb{N}$ and there is are two parameters: $p$ and $r$. The distribution is denoted as follows.

$$X \sim \mathrm{NB}\left(p, r\right)$$

- Before the $r$-th success at $x$, one needs $r - 1$ successes: the p.m.f. must account for all ways these successes can occur.

$$f_X\left(x; p, r\right) = \binom{x-1}{r-1} p^r \left(1 - p\right)^{x-r}$$

**Observation 1**

The geometric distribution is a special case of the negative binomial distribution, with $r = 1$; thus it is denoted as $X \sim \mathrm{NB}\left(p, 1\right)$

# The negative binomial distribution (2/2)

- One could also look at the number of **failures** $Y = X - r$:

$$f_Y(y; p, r) = \binom{r + y - 1}{y} p^r (1-p)^y$$
$$= (-1)^y \binom{-r}{y} p^r (1-p)^y$$

  whence the name “negative” binomial.

- The m.g.f. is defined for $t < -\log(1-p)$:

$$M_X(t; p, r) = \left( \frac{p \exp(t)}{1 - (1-p)\exp(t)} \right)^r$$

- …and one can obtain the following mean and variance.

$$\mathbb{E}[X] = \frac{r}{p}$$
$$\mathbb{V}\text{ar}[X] = \frac{r(1-p)}{p^2}$$

# The Poisson distribution (1/6)

- The Poisson distribution is an important distribution also connected to Bernoulli trials, albeit indirectly.
- Its support is $\mathbb{X} = \mathbb{N}_0 = \{0, 1, 2, \dots\}$.
- It has one parameter $\lambda \geq 0$, called **intensity**.
- The Poisson distribution is commonly denoted as:

$$X \sim \text{Pois}(\lambda)$$

- Its p.m.f., which is as follows, is not too intuitive.

$$f_X(x; \lambda) = \frac{\exp(-\lambda) \cdot \lambda^x}{x!}$$

# The Poisson distribution (2/6)

- The Poisson's c.d.f. is:

$$F_X(x;\lambda) = \exp(-\lambda) \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!}$$

and note that it sums to 1 over the support:

$$\begin{aligned} \lim_{M \to \infty} F_X(M;\lambda) &= \exp(-\lambda) \cdot \lim_{M \to \infty} \sum_{x=0}^{M} \frac{\lambda^x}{x!} \\ &= \exp(-\lambda) \cdot \exp(\lambda) \\ &= 1 \end{aligned}$$

(the Taylor expansion of exponential functions is used).

- An important property of Poisson distributions is that they **approximate** a binomial distribution when $p$ is **small** and $\lambda = pn$ (a demonstration follows).

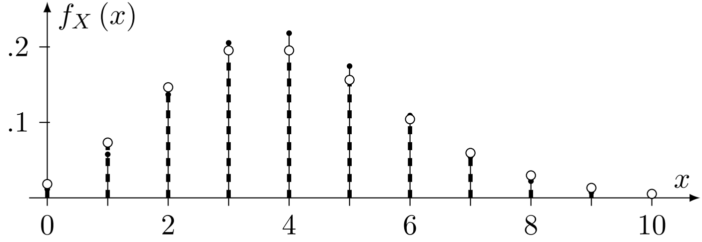
# The Poisson distribution (3/6)

Start from a binomial, fix $\lambda = pn$ and let $n \to \infty$ ($p \to 0$).

$$\begin{aligned}
\lim_{n\to\infty} f_X(x; p, n) &= \lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{n\to\infty} \frac{n!}{x!\,(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{\lambda^x}{x!} \lim_{n\to\infty} \underbrace{\left(\frac{\prod_{k=1}^{x}(n-k+1)}{n^x}\right)}_{\to 1} \cdot \\
&\quad \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{n}}_{\to \exp(-\lambda)} \\
&= \frac{\lambda^x}{x!} \exp(-\lambda) \\
&= f_X(x; \lambda)
\end{aligned}$$

# The Poisson distribution (4/6)

Binomial vs. Poisson comparison with $n = 20$, $p = 0.2$, $\lambda = 4$




- Binomial probabilities are denoted with solid thin lines, smaller full points; Poisson probabilities are denoted with dashed thicker lines, larger hollow points. Probabilities for $x > 10$ are negligible.

# The Poisson distribution (5/6)

It is now easier to interpret the Poisson distribution and its use.

- A model for the number of "occurrences" of some kind of event in a well-defined interval/space.

- Examples: phone calls, emails, holes in a piece of fabric.

- The occurrences of interest happen independently...

- ...and they are all equally likely.

- The larger the interval under examination, the higher the number of occurrences one can expect.

# The Poisson distribution (6/6)

- The Poisson's m.g.f. obtains by another Taylor expansion:

$$\begin{aligned} M_X(t;p) &= \lim_{M\to\infty} \sum_{x=0}^{M} \exp(tx) \cdot \frac{\exp(-\lambda) \cdot \lambda^x}{x!} \\ &= \exp(-\lambda) \cdot \lim_{M\to\infty} \sum_{x=0}^{M} \frac{[\lambda \cdot \exp(t)]^x}{x!} \\ &= \exp(-\lambda) \cdot \exp(\lambda \cdot \exp(t)) \\ &= \exp(\lambda (\exp(t) - 1)) \end{aligned}$$

- ...thus the key moments are both equal to:

$$\begin{aligned} \mathbb{E}[X] &= \lambda \\ \mathbb{V}\text{ar}[X] &= \lambda \end{aligned}$$

motivating the name "intensity" for $\lambda$.

# The uniform discrete distribution (1/2)

- Not all discrete distributions are based on Bernoulli trials! Suppose the probability is equal over entire support:

$$f_X(x; N) = \frac{1}{N}$$

for any $\mathbb{X}$ with $|\mathbb{X}| = N$: a *uniform discrete* distribution.

- If $\mathbb{X}$ are the integers between $a$ and $b$ (with $b - a = N - 1$):

$$X \sim \mathcal{U}\{a, b\}$$

- ...and the c.d.f. can be expressed as follows.

$$F_X(x; a, b) = \frac{\lfloor x \rfloor - a + 1}{b - a + 1} \cdot \mathbb{1}[a \leq x \leq b] + \mathbb{1}[b < x]$$

# The uniform discrete distribution (2/2)

- If $X \sim \mathcal{U}\{a, b\}$, the m.g.f. of $X$ is:

$$M_X(t; a, b) = \frac{\exp(at) - \exp((b+1)t)}{N(1 - \exp(t))}$$

- …and thus the mean and variance of $X$ are as follows.

$$\mathbb{E}[X] = \frac{a+b}{2}$$
$$\mathbb{V}\text{ar}[X] = \frac{N^2 - 1}{12}$$

It is perhaps easier to calculate these moments without the use of the m.g.f. here.

# Hypergeometric distribution (1/3)

- A Bernoulli trial, when repeated, is equivalent to the *urn experiment with replacement* (an urn contains balls of two kinds; when extracted, a ball is re-insterted in the urn).

- What if there is *no replacement* at every repetition? As in the binomial, interest falls on the number of **successes** $X$.

- Note: successes becomes less likely the more are obtained. For example, if $K$ balls can potentially return "success" at the first trial, these "good" balls decrease at every trial.

- The resulting *hypergeometric* distribution has parameters: $N$ (total possible occurrences/balls), $K$ (occurrences/balls that can deliver a success) and $n$ (number of trials).

- The distribution is denoted as follows.

$$X \sim \mathcal{H}(N, K, n)$$

# Hypergeometric distribution (2/3)

- The support of the hypergeometric distribution is:

$$\mathbb{X} = \{\max(0, n + K - N), \ldots, \min(n, K)\}$$

since it must be $x \geq 0$, $x \leq n$, $x \leq K$ and $n - x \leq N - K$.

- The p.m.f. (from which the c.d.f. can be derived) is:

$$f_X(x; N, K, n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

since there are $\binom{N}{n}$ ways to arrange **all** $n$ occurrences, $\binom{K}{x}$ ways to arrange the occurrences that deliver **success**, and $\binom{N-K}{n-x}$ to arrange the occurrences that deliver **failure**; all occurrences are equally likely.

# Hypergeometric distribution (3/3)

- Manipulating this distribution (e.g. to obtain the m.g.f.) is difficult due to the combinatorics involved.

- For example, the mean is calculated (for $Y = X - 1$) as:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in \mathbb{X}} x \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \\
&= \sum_{x \in \mathbb{X}} \frac{K\binom{K-1}{x-1}\binom{N-K}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} \\
&= n\frac{K}{N} \sum_{y \in \mathbb{Y}} \frac{\binom{K-1}{y}\binom{(N-1)-(K-1)}{n-1-y}}{\binom{N-1}{n-1}} \\
&= n\frac{K}{N}
\end{aligned}$$

- ...while the variance is $\mathbb{V}\mathrm{ar}[X] = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$.

# Moving to continuous distributions

- The analysis now moves to **continuous** distributions.
- It helps to begin from **location-scale** distribution families: these are characterized by two particular parameters.
- The **location** parameter determines the overall position of the distribution on $\mathbb{R}$ (usually associated with the mean). It is usually denoted by $\mu$.
- The **scale** parameter determines the overall "spread" of the distribution on $\mathbb{R}$ (usually associated with the variance). It is usually denoted by $\sigma$ (or $\sigma^2$), with $\sigma > 0$.
- Other distributions overviewed later may have other kinds of parameters that affect their overall **shape.**

# Location and scale

## Definition 1

**Location and scale families.** Let $f_Z(z)$ be a probability density function associated with some random variable $Z$. For any $\mu \in \mathbb{R}$ and any $\sigma \in \mathbb{R}_{++}$, the family of probability density functions of the form

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$$

for a generic random variable $X$ is called the **location-scale family** with **standard probability density function** $f_Z(z)$; $\mu$ is called the **location parameter** while $\sigma$ is called the **scale parameter**.

- Note: this implies $Z = (X - \mu)/\sigma$.
- Conversely, it is $X = \sigma Z + \mu$.

# Standardization of densities

## Theorem 1

**Standardization of densities.** *Let* $f(\cdot)$ *be any probability density function,* $\mu \in \mathbb{R}$ *and* $\sigma \in \mathbb{R}_{++}$*. Then, a random variable* $X$ *follows a probability distribution with density function:*

$$f_X(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

*if and only if there exists a continuous random variable* $Z$ *whose probability density function is* $f_Z(z) = f(z)$ *and* $X = \sigma Z + \mu$*.*

## Proof.

Necessity is shown through a density transformation with $X = g(Z)$: $g(Z) = \sigma Z + \mu$ (a monotone transformation), $g^{-1}(x) = (x-\mu)/\sigma$ and $\left|\frac{d}{dx} g^{-1}(x)\right| = 1/\sigma$. Sufficiency is shown by the converse exercise: define $Z = g(X) = (X-\mu)/\sigma$ – again a monotone transformation – with $g^{-1}(z) = \sigma z + \mu$, $\left|\frac{d}{dz} g^{-1}(z)\right| = \sigma$ and again one can extend the theorem for monotone transformations of density functions. □

# Implications of standardization

- A **standard** density function (with associated distribution) exists for every location-scale family.

- Given that all the distributions in a location-scale family are all linked through a linear transformation, their mean, variance, other moments and moment generating functions are related via simple functions.

- All **probabilities** that are specific to a distribution from a location-scale family can be expressed with reference to the standard distribution.

$$\begin{aligned}\mathbb{P}(a \leq X \leq b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)\end{aligned}$$

# The normal distribution (1/5)

- The *queen* of continuous distributions, a.k.a. "Gaussian."

- Given parameters $\mu$ and $\sigma^2$, it is indicated as:

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

  where its *standard* version is $Z \sim \mathcal{N}(0, 1)$.

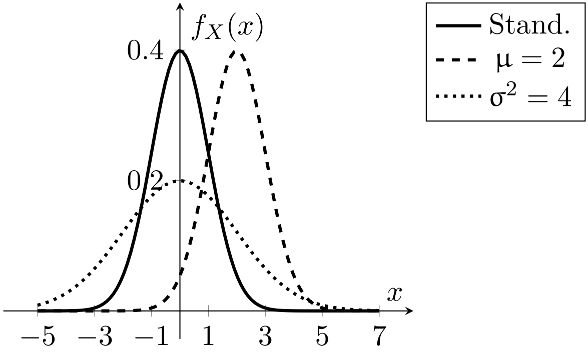
- Its support is $\mathbb{X} = \mathbb{R}$; the p.d.f. is:

$$f_X(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- ... while its c.d.f. obtains by integrating the density.

$$F_X(x; \mu, \sigma^2) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

# The normal distribution (2/5)



# The normal distribution (3/5)

- To show that the density integrates to 1, one can focus on the standard density, and specifically on half its support.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz = 1 \Leftrightarrow \int_{0}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz = \sqrt{\frac{\pi}{2}}$$

- The derivation is somewhat tedious.

$$\begin{aligned}
\left(\int_0^{\infty} \exp\left(-\frac{z^2}{2}\right) dz\right)^2 &= \left(\int_0^{\infty} \exp\left(-\frac{t^2}{2}\right) dt\right)\left(\int_0^{\infty} \exp\left(-\frac{u^2}{2}\right) du\right) \\
&= \int_0^{\infty}\int_0^{\infty} \exp\left(-\frac{t^2+u^2}{2}\right) dt du \\
&= \int_0^{\infty}\int_0^{\frac{\pi}{2}} r \cdot \exp\left(-\frac{r^2}{2}\right) d\theta dr \\
&= \frac{\pi}{2}\int_0^{\infty} r \cdot \exp\left(-\frac{r^2}{2}\right) dr \\
&= \frac{\pi}{2}\left[-\exp\left(-\frac{r^2}{2}\right)\Big|_0^{\infty}\right] \\
&= \frac{\pi}{2}
\end{aligned}$$

# The normal distribution (4/5)

- Obtaining the standard's m.g.f. is easier:

$$
\begin{aligned}
M_Z(t) &= \int_{-\infty}^{+\infty} \exp(tz) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2 - 2zt + t^2 - t^2}{2}\right) dz \\
&= \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-t)^2}{2}\right) dz \\
&= \exp\left(\frac{t^2}{2}\right)
\end{aligned}
$$

- hence, in the general case it is as follows.

$$
M_X\left(t; \mu, \sigma^2\right) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)
$$

# The normal distribution (5/5)

- The key moments of the normal distribution are:

$$\begin{aligned} \mathbb{E}[X] &= \mu \\ \mathbb{V}\mathrm{ar}[X] &= \sigma^2 \\ \mathbb{S}\mathrm{kew}[X] &= 0 \\ \mathbb{K}\mathrm{urt}[X] &= 3 \end{aligned}$$

  mean and variance coincide with $\mu$, $\sigma^2$; the constant $\mathbb{K}\mathrm{urt}$ is a reference point ($\mathbb{K}\mathrm{urt}[X] - 3$ is called "excess kurtosis").

- There is an alternative parametrization of the distribution:

$$f_X(x; \mu, \phi^2) = \sqrt{\frac{\phi^2}{2\pi}} \exp\left(-\frac{\phi^2 (x-\mu)^2}{2}\right)$$

  where $\phi^2 = \sigma^{-2}$ is called the *precision parameter*.

# The lognormal distribution (1/3)

- The *log*normal distribution obtains from the transformation $Y = \exp(X)$ where $X \sim \mathcal{N}(\mu, \sigma^2)$. Thus, the support of $Y$ is $\mathbb{Y} = \mathbb{R}_{++}$ but the parameters are still the normal's.

- From $X = \log(Y)$, the name and notation follow.

$$\log(Y) \sim \mathcal{N}(\mu, \sigma^2)$$

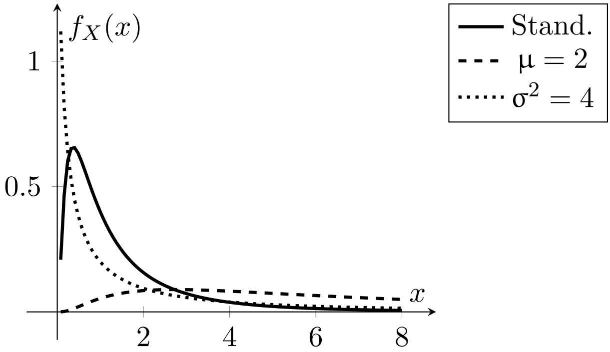
- Its p.d.f. is:

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right)$$

- ... while its c.d.f. obtains by integrating the density.

$$F_Y(y; \mu, \sigma^2) = \int_0^y \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{t} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right) dt$$

# The lognormal distribution (2/3)



# The lognormal distribution (3/3)

- The distribution lacks a m.g.f. but:

$$\mathbb{E}\left[Y^r\right] = \mathbb{E}\left[(\exp(X))^r\right] = \mathbb{E}\left[\exp(Xr)\right] = \exp\left(\mu r + \frac{\sigma^2 r^2}{2}\right)$$

- …which makes calculating moments easy:

$$\mathbb{E}\left[Y\right] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$\mathbb{V}\text{ar}\left[Y\right] = \left[\exp\left(\sigma^2\right) - 1\right] \exp\left(2\mu + \sigma^2\right)$$

- …while the skewness depends on $\sigma^2$ and is always positive.

$$\mathbb{S}\text{kew}\left[Y\right] = \left[\exp\left(\sigma^2\right) + 2\right] \cdot \sqrt{\exp(\sigma^2) - 1} > 0$$

# The logistic distribution (1/4)

- The logistic distribution has support $\mathbb{X} = \mathbb{R}$, parameters $\mu$ and $\sigma$, and a "bell shape" similar to the normal case.

- It can be denoted as follows.

$$X \sim \text{Logistic}(\mu, \sigma)$$

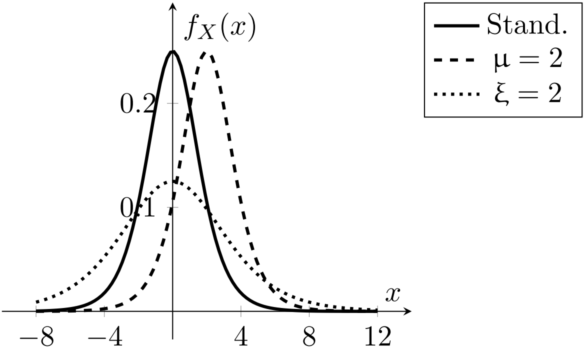
- Its p.d.f. is:

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \left[1 + \exp\left(-\frac{x-\mu}{\sigma}\right)\right]^{-2}$$

- ...while its c.d.f. is simpler to read.

$$F_X(x; \mu, \sigma) = \left[1 + \exp\left(-\frac{x-\mu}{\sigma}\right)\right]^{-1}$$

# The logistic distribution (2/4)



# The logistic distribution (3/4)

- The m.g.f. of the standard logistic is obtained as:

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} \exp(tz) \frac{\exp(-z)}{(1+\exp(-z))^2} dz \\ &= \int_0^1 u^t (1-u)^{-t} du \\ &= \mathrm{B}(1+t, 1-t) \end{aligned}$$

where $u = \frac{1}{1+\exp(-z)}$; observe that here $\frac{du}{dz} = \frac{\exp(-z)}{(1+\exp(-z))^2}$.

- Here $\mathrm{B}(a, b)$ for $a, b > 0$ denotes the **Beta function**:

$$\mathrm{B}(a, b) \equiv \int_0^1 u^{a-1} (1-u)^{b-1} du$$

- In the general case the m.g.f. is as follows.

$$M_X(t; \mu, \sigma) = \exp(\mu t) \cdot \mathrm{B}(1 - \xi t, 1 + \xi t)$$

# The logistic distribution (4/4)

- By the properties of the Beta function one can show that:

$$\mathbb{E}[X] = \mu$$
$$\mathbb{V}\text{ar}[X] = \frac{\sigma^2 \pi^2}{3}$$
$$\mathbb{S}\text{kew}[X] = 0$$
$$\mathbb{K}\text{urt}[X] = \frac{21}{5}$$

  observe the excess kurtosis!

- An obvious reparametrization of the logistic is $\sigma^* = \frac{\sqrt{3}}{\pi}\sigma$.

- The logistic has an important practical advantage – among others: a closed form **quantile function**!

$$Q_X(p; \mu, \sigma) = \mu + \sigma \log\left(\frac{p}{1-p}\right) \quad \text{for } p \in (0, 1)$$

# The Cauchy distribution (1/3)

- The Cauchy distribution has support $\mathbb{X} = \mathbb{R}$, parameters $\mu$ and $\sigma$, and a “bell shape” similar to the normal case.

- It can be denoted as follows.

$$X \sim \text{Cauchy}(\mu, \sigma)$$

- Its p.d.f. is:

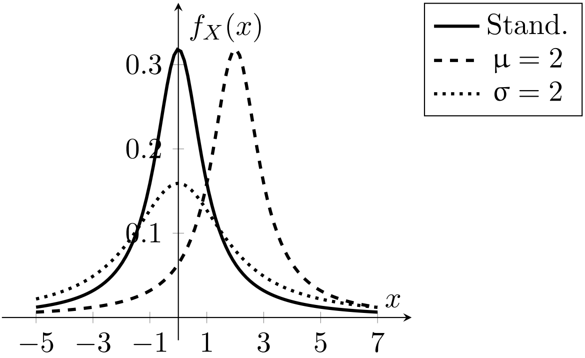
$$f_X(x; \mu, \sigma) = \frac{1}{\pi\sigma}\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-1}$$

- ...while its c.d.f. also has a closed form expression...

$$F_X(x; \mu, \sigma) = \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}$$

- ...which is invertible: $Q_X(p; \mu, \sigma) = \mu + \sigma\tan\left(\pi\left(p - \frac{1}{2}\right)\right)$!

# The Cauchy distribution (2/3)



# The Cauchy distribution (3/3)

- The Cauchy distribution is notorious for **lacking defined moments**. Consider its standard version's mean:

$$\mathbb{E}[Z] = \int_{-\infty}^{0} \frac{1}{\pi} \frac{z}{1+z^2}\, dz + \int_{0}^{+\infty} \frac{1}{\pi} \frac{z}{1+z^2}\, dz$$

these two halves are symmetric. But, take the latter:

$$\int_{0}^{+\infty} \frac{z}{1+z^2}\, dz = \lim_{M\to\infty} \left. \frac{\log\left(1+z^2\right)}{2} \right|_{0}^{M} = \infty$$

the integrals diverge! The mean **cannot be defined**.

- The Cauchy lacks a m.g.f. but, like all distributions, has a **characteristic function** which is discontinuous at $t = 0$!

$$\varphi_X(t; \mu, \sigma) = \exp\left(i\mu t - \sigma |t|\right)$$

# The Laplace distribution (1/4)

- The Laplace distribution has support $\mathbb{X} = \mathbb{R}$, parameters $\mu$ and $\sigma$, and a characteristic "tent shape."

- It is also known as "double exponential" (for reasons to be clarified later) and can be denoted as follows.

$$X \sim \text{Laplace}\left(\mu, \sigma\right)$$

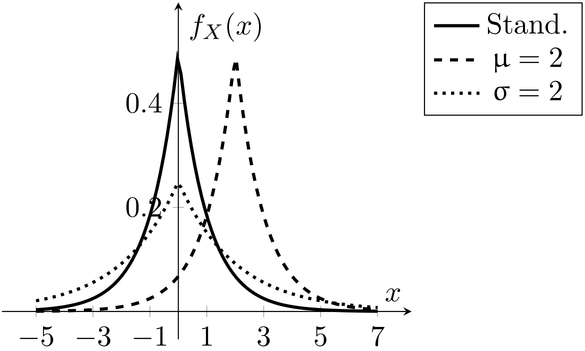
- Its p.d.f. features an absolute value:

$$f_X\left(x; \mu, \sigma\right) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right)$$

- ...thus, its c.d.f. depends on the value of $x$.

$$F_X\left(x; \mu, \sigma\right) = \begin{cases} \frac{1}{2} \exp\left(\frac{x-\mu}{\sigma}\right) & \text{if } x < \mu \\ 1 - \frac{1}{2} \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{if } x \geq \mu \end{cases}$$

# The Laplace distribution (2/4)



# The Laplace distribution (3/4)

- As usual, it is easier to calculate the standard m.g.f. first:

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{+\infty} \frac{1}{2} \exp\left(tz - |z|\right) dz \\ &= \frac{1}{2} \int_{-\infty}^{0} \exp\left(\left(1+t\right) z\right) dz + \\ &\quad + \frac{1}{2} \int_{0}^{+\infty} \exp\left(-\left(1-t\right) z\right) dz \\ &= \frac{1}{2} \left[ \frac{1}{1+t} + \frac{1}{1-t} \right] \\ &= \frac{1}{1-t^2} \end{aligned}$$

- ...so to easily generalize it (note: only for $|t| < \sigma^{-1}$).

$$M_X(t; \mu, \sigma) = \frac{\exp(\mu t)}{1 - \sigma^2 t^2}$$

# The Laplace distribution (4/4)

- The mean and variance are as follows.

$$\mathbb{E}\left[X\right] = \mu$$
$$\mathbb{V}\text{ar}\left[X\right] = 2\sigma^2$$

- Like the logistic and Cauchy, this distribution possesses an explicit quantile function.

$$Q_X\left(p; \mu, \sigma\right) = \begin{cases} \mu + \sigma \log\left(2p\right) & \text{if } p \in \left(0, \frac{1}{2}\right] \\ \mu - \sigma \log\left(2 - 2p\right) & \text{if } p \in \left[\frac{1}{2}, 1\right) \end{cases}$$

- The Laplace distribution has some limited applications in the social sciences; these include modeling *growth rates* of certain populations (e.g. firms).

# Beyond location-scale families

- The rest of the analysis concerns continuous distributions that do not belong to a strict location-scale family.

- The parameters of these distributions may be pure **shape** parameters, and/or determine the **support.**

- It is useful to always specify the **range** of admissible values for these parameters.

- There is plenty of **relationships** within and between these distributions.

- The analysis ends with **extreme value** distributions, that feature **three** parameters: for location, scale and shape.

# The uniform distribution (1/2)

- Uniform distributions have **bounded support**: $\mathbb{X} = [a, b]$ with $a \leq b$, but the interval may as well be *open*.

- Here $a$ and $b$ are effectively **parameters**. The notation is:

$$X \sim \mathcal{U}(a, b)$$

(parentheses and not brackets, unlike the discrete uniform).

- The p.d.f. makes use of indicator functions:

$$f_X(x; a, b) = \frac{1}{b-a} \cdot \mathbb{1}[x \in (a, b)]$$

- ...and the c.d.f. too.

$$F_X(x; a, b) = \frac{x-a}{b-a} \cdot \mathbb{1}[x \in (a, b)] + \mathbb{1}[x \in [b, \infty)]$$

# The uniform distribution (2/2)

- If $X \sim \mathcal{U}(a, b)$, the m.g.f. of $X$ is:

$$M_X(t; a, b) = \begin{cases} \frac{1}{t(b-a)} \left[\exp(bt) - \exp(at)\right] & \text{if } t \neq 0 \\ 1 & \text{if } t = 0 \end{cases}$$

- ... and thus the mean and variance of $X$ are as follows.

$$\mathbb{E}[X] = \frac{a+b}{2}$$

$$\mathbb{V}\text{ar}[X] = \frac{(b-a)^2}{12}$$

It is perhaps easier to calculate these moments without the use of the m.g.f. here.

- The analysis does not change if the support is *open*.

# The Beta distribution (1/5)

- Beta distributions are general distributions with **bounded support**. Focus for now on $\mathbb{X} = [0,1]$, or $\mathbb{X} = (0,1)$.

- Here $\alpha > 0$ and $\beta > 0$ are the **parameters**; the notation is:

$$X \sim \text{Beta}(\alpha, \beta)$$

- which is motivated by a p.d.f. expressed through the Beta function (a normalizing factor):

$$f_X(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta)}$$

- ...just like the c.d.f.!

$$F_X(x; \alpha, \beta) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,dt}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} = \frac{\mathrm{B}(x; \alpha, \beta)}{\mathrm{B}(\alpha, \beta)}$$

# The Beta distribution (2/5)

- In the expression for the c.d.f., $\mathrm{B}(x; \alpha, \beta)$ is the so-called **lower incomplete Beta function**; for any positive $a, b$:

$$\mathrm{B}(x; a, b) \equiv \int_0^x u^{a-1} (1-u)^{b-1} \, du.$$

- The so-called **Gamma function** $\Gamma(c)$, for $c > 0$:

$$\Gamma(c) = \int_0^\infty u^{c-1} \exp(-u) \, du$$

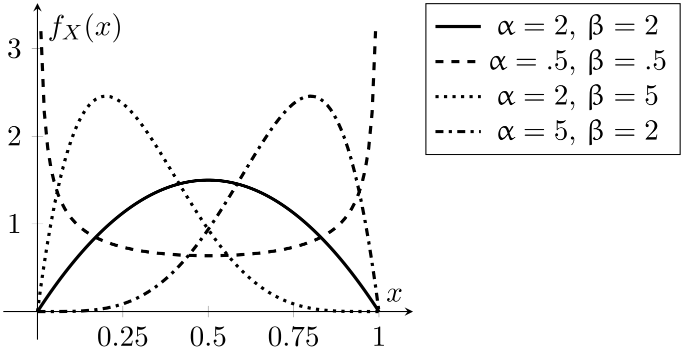
is related to the Beta function:

$$\mathrm{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

hence the p.d.f. can be alternatively written as follows.

$$f_X(x; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1} (1-x)^{\beta-1}$$

# The Beta distribution (3/5)




**Observation 2**

$X \sim \text{Beta}\,(1, 1)$ is equivalent to $X \sim \mathcal{U}\,(0, 1)$.

# The Beta distribution (4/5)

- The Beta's m.g.f. is difficult to obtain:

$$M_X\left(t;\alpha,\beta\right)=1+\sum_{q}^{\infty}\left(\prod_{k=0}^{q-1}\frac{\alpha+k}{\alpha+\beta+k}\right)\frac{t^q}{q!}$$

- ... and uncentered moments are best calculated directly!

$$\begin{aligned}\mathbb{E}\left[X^r\right]&=\frac{1}{\mathrm{B}\left(\alpha,\beta\right)}\int_0^1 x^{r+\alpha-1}\left(1-x\right)^{\beta-1}dx\\&=\frac{\mathrm{B}\left(r+\alpha,\beta\right)}{\mathrm{B}\left(\alpha,\beta\right)}\\&=\frac{\Gamma\left(r+\alpha\right)\Gamma\left(\beta\right)}{\Gamma\left(r+\alpha+\beta\right)}\cdot\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\\&=\frac{\Gamma\left(r+\alpha\right)\Gamma\left(\alpha+\beta\right)}{\Gamma\left(r+\alpha+\beta\right)\Gamma\left(\alpha\right)}\end{aligned}$$

# The Beta distribution (5/5)

- A Gamma function's property greatly helps calculations:

$$\Gamma(c) = (c-1) \cdot \Gamma(c-1)$$

and the key moments are obtained as follows.

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$$

$$\mathbb{V}\mathrm{ar}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- This analysis generalizes to **any connected segment** of $\mathbb{R}$ for support. For $\mathbb{X} = [a, b]$ or $\mathbb{X} = (a, b)$:

$$f_X(x; \alpha, \beta, a, b) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{\mathrm{B}(\alpha, \beta) \cdot (b-a)^{\alpha+\beta-1}}$$

is a *nonstandard* Beta (where if $\alpha = \beta = 1$, $X \sim \mathcal{U}(a, b)$).

# The exponential distribution (1/4)

- The exponential distributions are simple distributions with support on the set of nonnegative real numbers, $\mathbb{X} = \mathbb{R}_+$.

- There is **one** parameter $\lambda > 0$ (note: often reparametrized as $\beta = \lambda^{-1}$) and the notation is the following.

$$X \sim \text{Exp}(\lambda)$$

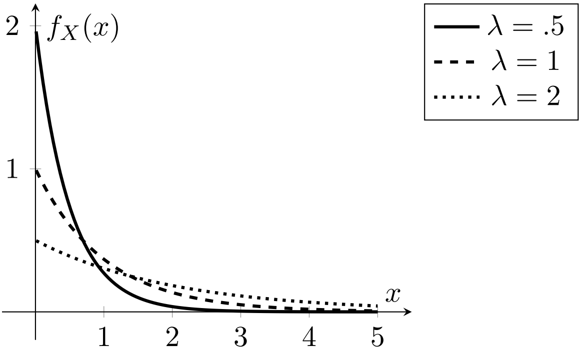
- The name comes from the functional form of the p.d.f.:

$$f_X(x; \lambda) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right)$$

- …as well as that of the c.d.f. (unsurprisingly).

$$F_X(x; \lambda) = 1 - \exp\left(-\frac{x}{\lambda}\right)$$

# The exponential distribution (2/4)



# The exponential distribution (3/4)

- It is easy to obtain the m.g.f. (recall the case with $\lambda = 1$), however it only exists for $t < \lambda^{-1}$:

$$M_X(t; \lambda) = \frac{1}{1 - \lambda t}$$

- and the key moments are as follows.

$$\mathbb{E}[X] = \lambda$$
$$\mathbb{V}\text{ar}[X] = \lambda^2$$

- This distribution is the continuous analog of the *geometric* distribution. They both share the *memoryless* property:

$$\mathbb{P}(X > s | X > t) = \mathbb{P}(X > s - t)$$

(the derivation is similar). The exponential distribution is used to model (continuous) *waiting times*.

# The exponential distribution (4/4)

**Observation 3**
If $X \sim \mathcal{U}(0,1)$ and $Y = -\lambda \log(X)$ it is $Y \sim \operatorname{Exp}(\lambda)$.

**Observation 4**
If $X \sim \operatorname{Exp}(\lambda)$ and $Y = \exp(-X)$ it is $Y \sim \operatorname{Beta}\left(\frac{1}{\lambda}, 1\right)$.

**Observation 5**
If $X \sim \operatorname{Laplace}(\mu, \sigma)$ and $Y = |X - \mu|$ it is $Y \sim \operatorname{Exp}(\sigma)$, whence the name "double exponential" for the Laplace.

**Observation 6**
If $X \sim \operatorname{Exp}(1)$ and

$$Y = \mu - \sigma \log\left(\frac{\exp(-X)}{1 - \exp(-X)}\right)$$

it is $Y \sim \operatorname{Logistic}(\mu, \sigma)$. The standard logistic models the *odds ratio* of exponential events.

# The Gamma distribution (1/4)

- Gamma distributions have support upon the set of positive real numbers $\mathbb{X} = \mathbb{R}_{++}$ ($X = 0$ may be included at will).

- There are **two** parameters $\alpha > 0$ and $\beta > 0$ (the latter can be reparametrized as $\theta = \beta^{-1}$); two notations coexist.

$$X \sim \Gamma(\alpha, \beta) \quad \& \quad X \sim \text{Gamma}(\alpha, \beta)$$

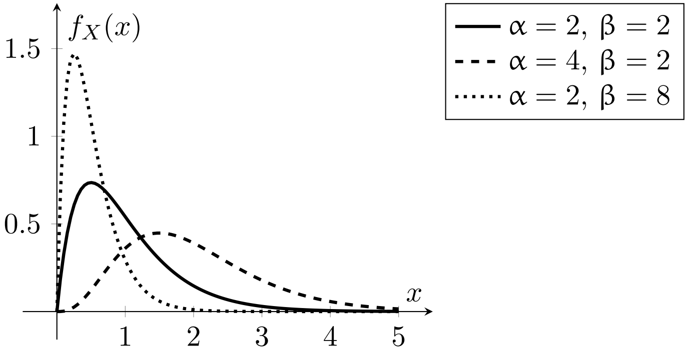
- As expected, the Gamma function normalizes the p.d.f.:

$$f_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} \exp(-\beta x)$$

- ...so the c.d.f. can be expressed via the **lower incomplete Gamma function**: $\gamma(a, b) = \int_0^b u^{a-1} \exp(-u)\, du$.

$$F_X(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \int_0^x \beta^{\alpha} t^{\alpha-1} \exp(-\beta t)\, dt = \frac{\gamma(\alpha, x)}{\Gamma(\alpha)}$$

# The Gamma distribution (2/4)




**Observation 7**

$X \sim \text{Gamma}\left(1, \frac{1}{\lambda}\right)$ is equivalent to $X \sim \text{Exp}\left(\lambda\right)$, that is, exponential distributions are all special cases of the Gamma family.

# The Gamma distribution (3/4)

- Uncentered moments are better calculated directly:

$$\mathbb{E}\left[X^r\right] = \frac{1}{\Gamma(\alpha)}\frac{1}{\beta^r}\int_0^{\infty} \beta^{r+\alpha} x^{r+\alpha-1} \exp(-\beta x)\, dx$$
$$= \frac{\Gamma(r+\alpha)}{\Gamma(\alpha)\beta^r}$$

  the integral, rescaled by $\Gamma(r+\alpha)$, is the p.d.f. of a Gamma distribution with parameters $r+\alpha$ and $\beta$.

- Using the property $\Gamma(c) = (c-1)\cdot\Gamma(c-1)$ again, the key moments are as follows.

$$\mathbb{E}[X] = \frac{\alpha}{\beta}$$
$$\mathbb{V}\text{ar}[X] = \frac{\alpha}{\beta^2}$$

# The Gamma distribution (4/4)

- Alternatively, one could have calculated the m.g.f. as:

$$\begin{aligned}
M_X(t;\alpha,\beta) &= \frac{1}{\Gamma(\alpha)}\int_0^\infty \exp(tx)\,\beta^\alpha x^{\alpha-1}\exp(-\beta x)\,dx \\
&= \frac{\beta^\alpha}{(\beta-t)^\alpha}\int_0^\infty \frac{(\beta-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-(\beta-t)\,x)\,dx \\
&= \left(\frac{\beta}{\beta-t}\right)^\alpha \\
&= \left(1-\frac{t}{\beta}\right)^{-\alpha}
\end{aligned}$$

  within the integral is a Gamma p.d.f. with parameters $\alpha$ and $\beta - t$. The m.g.f. is only defined for $t < \beta$!

- Gamma distributions have a wide range of applications for flexibly modeling phenomena with support on $\mathbb{X} = \mathbb{R}_+$.

# The Chi-squared distribution (1/3)

- Chi-squared distributions also have support upon the set of positive real numbers $\mathbb{X} = \mathbb{R}_{++}$ (with possibly $X = 0$).

- There is **one** parameter $\kappa > 0$; when $\kappa \in \mathbb{N}$ (integer), this is known as **degrees of freedom**. The notation is as follows.

$$X \sim \chi^2(\kappa) \quad \text{or} \quad X \sim \chi^2_\kappa$$

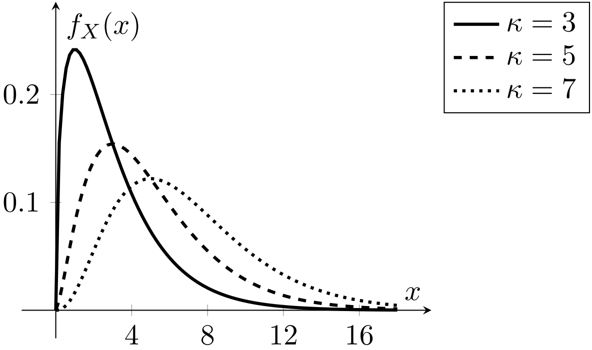
- The p.d.f. is normalized by the Gamma function:

$$f_X(x;\kappa) = \frac{1}{\Gamma\left(\frac{\kappa}{2}\right) \cdot 2^{\frac{\kappa}{2}}} x^{\frac{\kappa}{2}-1} \exp\left(-\frac{x}{2}\right)$$

  and so does the c.d.f. (not reported for brevity).

- It is obvious that this is a subfamily of the Gamma family. It is singled out because of its role in *statistical inference*.

# The Chi-squared distribution (2/3)




**Observation 8**

$X \sim \text{Gamma}\left(\frac{\kappa}{2}, \frac{1}{2}\right)$ is equivalent to $X \sim \chi^2(\kappa)$, that is, chi-squared distributions are all special cases of the Gamma family.

# The Chi-squared distribution (3/3)

- From the Gamma's analysis, the m.g.f. is (for $t < 0.5$):

$$M_X(t; \kappa) = (1 - 2t)^{-\frac{\kappa}{2}}$$

- ... while the key moments are as follows.

$$\mathbb{E}[X] = \kappa$$
$$\mathbb{V}\text{ar}[X] = 2\kappa$$

**Observation 9**
$X \sim \chi^2(2)$ is equivalent to $X \sim \text{Exp}\left(\frac{1}{2}\right)$.

**Observation 10**
If $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$, it is $Y \sim \chi^2(1)$.

# Snedecor's $\boldsymbol{F}$-distribution (1/3)

- Another family of distribution with support upon the set of positive real numbers $\mathbb{X} = \mathbb{R}_{++}$ (with possibly $X = 0$).

- There are **two** parameters $\nu_1 > 0$ and $\nu_2 > 0$, called paired **degrees of freedom** if integers. The notation is as follows.

$$X \sim \mathcal{F}\left(\nu_1, \nu_2\right) \quad \text{or} \quad X \sim \mathcal{F}_{\nu_1, \nu_2}$$

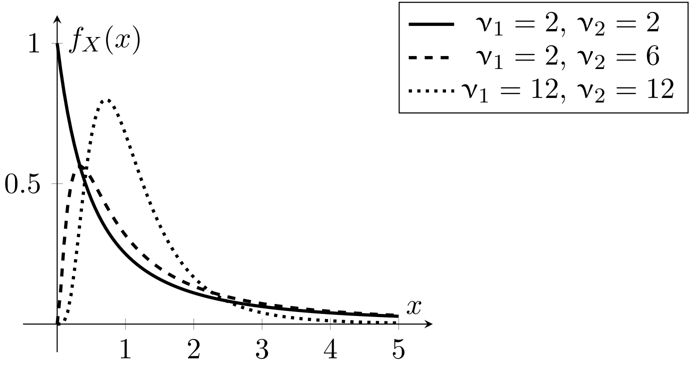
- The p.d.f. is normalized by the Beta function:

$$f_X\left(x; \nu_1, \nu_2\right) = \frac{x^{\frac{\nu_1}{2}-1}}{\mathrm{B}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \left(1 + \frac{\nu_1}{\nu_2} x\right)^{-\frac{\nu_1+\nu_2}{2}}$$

- ...thus the c.d.f. is best expressed via the incomplete Beta function.

$$F_X\left(x; \nu_1, \nu_2\right) = \frac{\mathrm{B}\left(x, \frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}{\mathrm{B}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)}$$

# Snedecor's *F*-distribution (2/3)




**Observation 11**
If $X \sim \mathcal{F}(\nu_1, \nu_2)$ and $Y = X^{-1}$, it is $Y \sim \mathcal{F}(\nu_2, \nu_1)$.

# Snedecor's *F*-distribution (3/3)

- The $F$-distribution lacks a m.g.f. and also its characteristic function is involved. Key moments are better obtained via direct integration.

$$\mathbb{E}[X] = \frac{\nu_2}{\nu_2 - 2}$$

$$\mathbb{V}\text{ar}[X] = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$$

- The $F$-distribution also plays a role in statistical inference.

**Observation 12**
If $X \sim \mathcal{F}(\nu_1, \nu_2)$ and $Y \sim \text{Beta}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)$, the random variables $X$ and $Y$ are related through the following reciprocal transformations.

$$Y = \frac{(\nu_1 X/\nu_2)}{(1 - \nu_1 X/\nu_2)} \qquad X = \frac{\nu_2 Y}{\nu_1(1 - Y)}$$

# Student's *t*-distribution (1/4)

- Back to a bell-shaped family with "full" support $\mathbb{X} = \mathbb{R}$!
- There is **one** parameter $\nu > 0$; when $\nu \in \mathbb{N}$ (integer), this is known as **degrees of freedom**. The notation is as follows.

$$X \sim \mathcal{T}(\nu) \quad \text{or} \quad X \sim \mathcal{T}_\nu$$

- The p.d.f. is normalized by the Beta function:

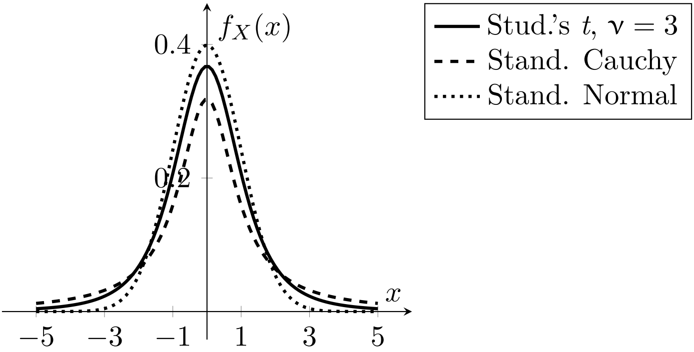
$$f_X(x;\nu) = \frac{1}{\mathrm{B}\left(\frac{1}{2}, \frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

or even by the Gamma, since $\mathrm{B}\left(\frac{1}{2}, \frac{\nu}{2}\right) = \Gamma\left(\frac{\nu}{2}\right) \Big/ \Gamma\left(\frac{\nu+1}{2}\right)$.

- The $t$-distribution's c.d.f. is perhaps best expressed through the incomplete Beta function.

$$F_X(x;\nu) = \begin{cases} \frac{1}{2}\mathrm{B}\left(\frac{\nu}{x^2+\nu}, \frac{1}{2}, \frac{\nu}{2}\right) & \text{if } x \le 0 \\ 1 - \frac{1}{2}\mathrm{B}\left(\frac{\nu}{x^2+\nu}, \frac{1}{2}, \frac{\nu}{2}\right) & \text{if } x > 0 \end{cases}$$

# Student's *t*-distribution (2/4)




**Observation 13**

$X \sim \mathcal{T}(1)$ is equivalent to $X \sim \text{Cauchy}(0, 1)$.

# Student's *t*-distribution (3/4)

- The $t$-distribution lacks a m.g.f., its characteristic function is involved, and moments of order $r \geq \nu$ are not defined.

$$\mathbb{E}\left[X^r\right] = \begin{cases} \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{\nu-r}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \sqrt{\frac{\nu^r}{\pi}} & \text{if } r \text{ is even, } 0 < r < \nu \\ 0 & \text{if } r \text{ is odd, } 0 < r < \nu \end{cases}$$

- Hence, key moments exist only for some values of $\nu$.

$$\begin{aligned} \mathbb{E}\left[X\right] &= 0 && \text{for } \nu > 1 \\ \mathbb{V}\text{ar}\left[X\right] &= \frac{\nu}{\nu-2} && \text{for } \nu > 2 \\ \mathbb{S}\text{kew}\left[X\right] &= 0 && \text{for } \nu > 3 \\ \mathbb{K}\text{urt}\left[X\right] &= \frac{3\nu-6}{\nu-4} && \text{for } \nu > 4 \end{aligned}$$

# Student's *t*-distribution (4/4)

**Observation 14**
If $X \sim \mathcal{T}(\nu)$ and $Y = X^2$, it is $Y \sim \mathcal{F}(1, \nu)$.

**Observation 15**
If $X \sim \mathcal{T}(\nu)$ and $Y = X^{-2}$, it is $Y \sim \mathcal{F}(\nu, 1)$.

- The *t*-distribution is central in statistical inference, largely because of some results that relate it to both the standard normal and the chi-squared distribution (Lectures 3, 4).

- But this is also due to its asymptotic relationship with the standard normal, which the *t*-distribution approximates as $\nu \to \infty$ (Lecture 6).

# The Pareto distribution (1/4)

- This distribution has a support which depends on one of its parameters: $\mathbb{X} = [\alpha, \infty)$, where $\alpha > 0$.

- There is also a **second** parameter: $\beta > 0$. The notation for a Pareto distribution is unsurprisingly as follows.

$$X \sim \text{Pareto}(\alpha, \beta)$$

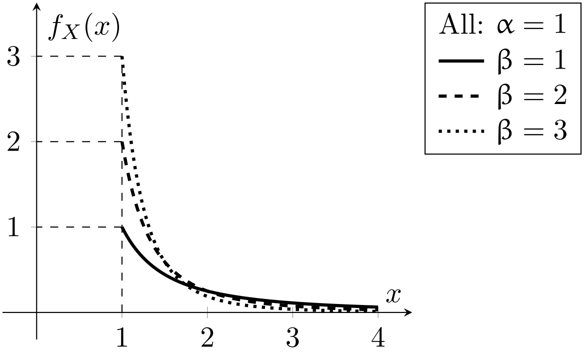
- The p.d.f. is (note the specification of the support):

$$f_X(x; \alpha, \beta) = \frac{\beta \alpha^{\beta}}{x^{\beta+1}} \quad \text{for } x \geq \alpha$$

- ...while the c.d.f. is as follows.

$$F_X(x; \alpha, \beta) = 1 - \left(\frac{\alpha}{x}\right)^{\beta} \quad \text{for } x \geq \alpha$$

# The Pareto distribution (2/4)




**Observation 16**

If $X \sim \text{Pareto}(\alpha, \beta)$ and $Y \sim \text{Exp}\left(\beta^{-1}\right)$, the two random variables are related through the two symmetric transformations $X = \alpha \exp(Y)$ and $Y = \log(X/\alpha)$.

# The Pareto distribution (3/4)

- The m.g.f. is expressed here through the **upper incomplete Gamma function**: $\Gamma(a,b) = \int_b^{\infty} u^{a-1} \exp(-u)\, du$.

$$M_X(x;\alpha,\beta) = \beta(-\alpha t)^{\beta} \cdot \Gamma(-\beta,-\alpha t)$$

- Key moments are best obtained via direct integration, but they exist only for some values of $\beta$. The mean is:

$$\mathbb{E}[X] = \begin{cases} \infty & \text{for } \beta \leq 1 \\ \dfrac{\alpha\beta}{\beta-1} & \text{for } \beta > 1 \end{cases}$$

while the variance is as follows.

$$\mathbb{V}\text{ar}[X] = \begin{cases} \infty & \text{for } \beta \leq 2 \\ \dfrac{\alpha^2\beta}{(\beta-1)^2(\beta-2)} & \text{for } \beta > 2 \end{cases}$$

# The Pareto distribution (4/4)

- Pareto distributions feature a so-called "fat tail:" extreme realizations of $X$ are relatively likely.
- They are also noteworthy for their **Power Law**: in logs the p.d.f. is conveniently linear.

$$\log f_X(x; \alpha, \beta) = \log\left(\beta\alpha^{\beta}\right) - (\beta + 1)\log x \quad \text{for } x \geq \alpha$$

- Their quantile function is also a simple expression.

$$Q_X(p; \alpha, \beta) = \alpha(1 - p)^{-\frac{1}{\beta}}$$

- There is a "generalized" family of Pareto distributions:

$$F_X(x; \beta, \gamma, \mu, \sigma) = 1 - \left[1 + \left(\frac{x - \mu}{\sigma}\right)^{\frac{1}{\gamma}}\right]^{-\beta} \quad \text{for } x \geq \mu$$

with $\mu \in \mathbb{R}$, $(\beta, \gamma, \sigma) \in \mathbb{R}^3_{++}$ and support $\mathbb{X} = [\mu, \infty)$.

# Generalized Extreme Value distributions (1/4)

- The family of Generalized Extreme Value distribution is a large one, and includes several subfamilies.

- It gets its name from its connection with the Extreme Value Theorem (Lecture 6). These distributions are **fat-tailed**.

- The family features **three parameters**: $\mu \in \mathbb{R}$ (**location**), $\sigma \in \mathbb{R}_{++}$ (**scale**), and $\xi \in \mathbb{R}$ (**shape**). A notation valid for the whole family is as follows.

$$X \sim \text{GEV}(\mu, \sigma, \xi)$$

- The support depends on the value of the shape parameter.

$$\mathbb{X} = \begin{cases} \left[\mu - \frac{\sigma}{\xi}, \infty\right) & \text{if } \xi > 0 \\ (-\infty, \infty) & \text{if } \xi = 0 \\ \left(-\infty, \mu - \frac{\sigma}{\xi}\right] & \text{if } \xi < 0 \end{cases}$$

# Generalized Extreme Value distributions (2/4)

- The p.d.f. also depends on the shape parameter $\xi$:

$$f_Z(z;\xi) = \begin{cases} \dfrac{\exp\left(-\left(1+\xi z\right)^{-\frac{1}{\xi}}\right)}{\left(1+\xi z\right)^{\frac{1}{\xi}+1}} & \text{for } \xi \neq 0 \text{ and } \xi z > -1 \\ \dfrac{\exp\left(-\exp\left(-z\right)\right)}{\exp\left(z\right)} & \text{for } \xi = 0 \end{cases}$$

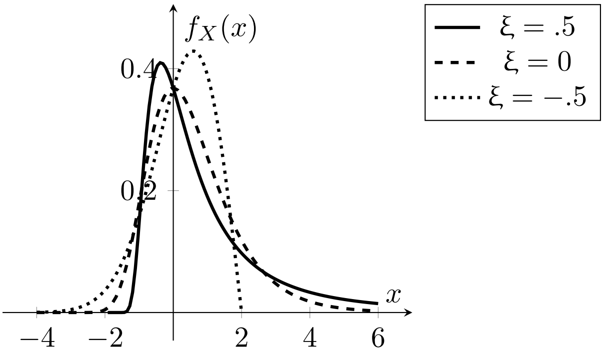
- …and so does the c.d.f.:

$$F_Z(z;\xi) = \begin{cases} \exp\left(-\left(1+\xi z\right)^{-\frac{1}{\xi}}\right) & \text{for } \xi \neq 0 \text{ and } \xi z > -1 \\ \exp\left(-\exp\left(-z\right)\right) & \text{for } \xi = 0 \end{cases}$$

- …and so does the quantile function.

$$Q_Z(p;\xi) = \begin{cases} \dfrac{\left(-\log\left(p\right)\right)^{-\xi} - 1}{\xi} & \text{for } \xi \neq 0 \\ -\log\left(-\log\left(p\right)\right) & \text{for } \xi = 0 \end{cases}$$

# Generalized Extreme Value distributions (3/4)




- Type I Extreme Value: $\xi = 0$ (Gumbel)
- Type II Extreme Value: $\xi > 0$ (Fréchet)
- Type III Extreme Value: $\xi < 0$ (reverse Weibull)

# Generalized Extreme Value distributions (4/4)

- The m.g.f. and characteristic functions are quite involved.
- Moments are better obtained via direct integration, but are defined for some values of $\xi$ only.
- The mean is given by:

$$\mathbb{E}[X] = \begin{cases} \mu + \frac{\sigma}{\xi}\left[\Gamma(1-\xi) - 1\right] & \text{if } \xi \neq 0, \xi < 1 \\ \mu + \sigma\gamma & \text{if } \xi = 0 \\ \infty & \text{if } \xi \geq 1 \end{cases}$$

- ... while the variance is as follows.

$$\mathbb{V}\text{ar}[X] = \begin{cases} \frac{\sigma^2}{\xi^2}\left[\Gamma(1-2\xi) - (\Gamma(1-\xi))^2\right] & \text{if } \xi \neq 0, \xi < \frac{1}{2} \\ \sigma^2 \frac{\pi^2}{6} & \text{if } \xi = 0 \\ \infty & \text{if } \xi \geq \frac{1}{2} \end{cases}$$

# The Gumbel (Type I GEV) distribution (1/2)

- The simplest GEV distributions (**Type I**) have $\xi = 0$ and only a location and scale parameter.

- Two alternative pieces of notation are used for them.

$$X \sim \text{EV1}(\mu, \sigma) \quad \& \quad X \sim \text{Gumbel}(\mu, \sigma)$$

- The p.d.f. is given by:

$$f_X(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$$

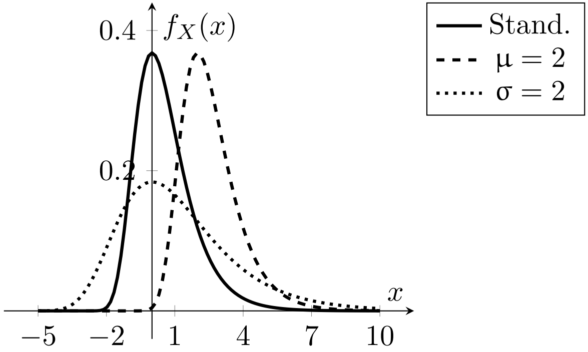
- ... the c.d.f. is:

$$F_X(x; \mu, \sigma) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right)$$

- ... while the quantile function is as follows.

$$Q_X(p; \mu, \sigma) = \mu - \sigma \log(-\log(p))$$

# The Gumbel (Type I GEV) distribution (2/2)



# The Fréchet (Type II GEV) distribution (1/2)

- The **Type II** GEV distributions are usually rephrased via $\alpha \equiv \xi^{-1} > 0$ & the transformation $Y = \sigma + \mu(1-\xi) + \xi X$.

- Two alternative pieces of notation are used for them.

$$Y \sim \text{EV2}(\alpha, \mu, \sigma) \quad \& \quad Y \sim \text{Frechet}(\alpha, \mu, \sigma)$$

- The p.d.f. is given by:

$$f_Y(y; \alpha, \mu, \sigma) = \frac{\alpha}{\sigma}\left(\frac{y-\mu}{\sigma}\right)^{-\alpha-1} \exp\left(-\left(\frac{y-\mu}{\sigma}\right)^{-\alpha}\right)$$

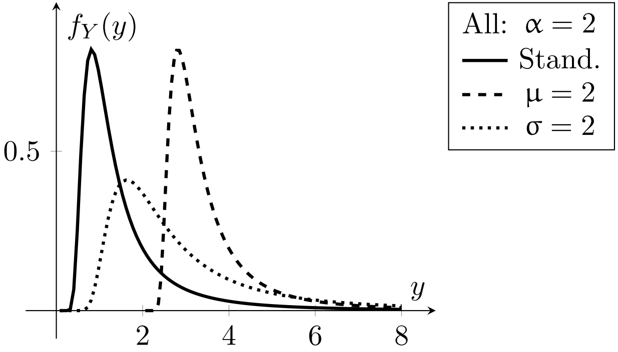
- ...the c.d.f. is:

$$F_Y(y; \alpha, \mu, \sigma) = \exp\left(-\left(\frac{y-\mu}{\sigma}\right)^{-\alpha}\right)$$

- ...while the quantile function is as follows.

$$Q_Y(p; \alpha, \mu, \sigma) = (-\log(p))^{\frac{1}{\alpha}}$$

# The Fréchet (Type II GEV) distribution (2/2)




- Recall that the support is $\mathbb{Y} = [\mu, \infty)$.

# The Weibull (Type III GEV) distribution (1/4)

- The **Type III** GEV distributions also feature $\alpha \equiv \xi^{-1} < 0$.
- Here, $Y$ is said to follow the **reverse Weibull** distribution. The symmetric $W = -Y$ follows the "traditional" **Weibull** distribution. Each has its own notation.

$$Y \sim \text{EV3}\left(\alpha, \mu, \sigma\right) \quad \& \quad W \sim \text{Weibull}\left(\alpha, \mu, \sigma\right)$$

- The p.d.f. of the *traditional* Weibull is:

$$f_W\left(w; \alpha, \mu, \sigma\right) = \frac{\alpha}{\sigma}\left(\frac{w-\mu}{\sigma}\right)^{\alpha-1} \exp\left(-\left(\frac{w-\mu}{\sigma}\right)^{\alpha}\right)$$

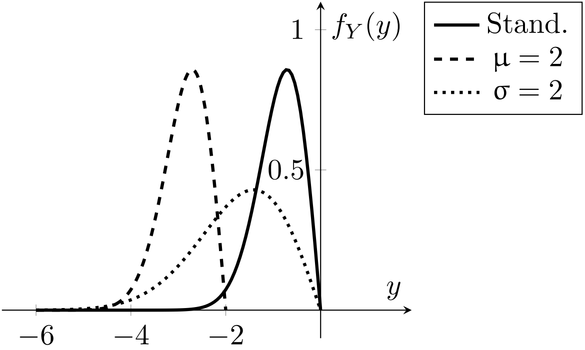
- …its c.d.f. is:

$$F_W\left(w; \alpha, \mu, \sigma\right) = 1 - \exp\left(-\left(\frac{w-\mu}{\sigma}\right)^{\alpha}\right)$$

- …while its quantile function is as follows.

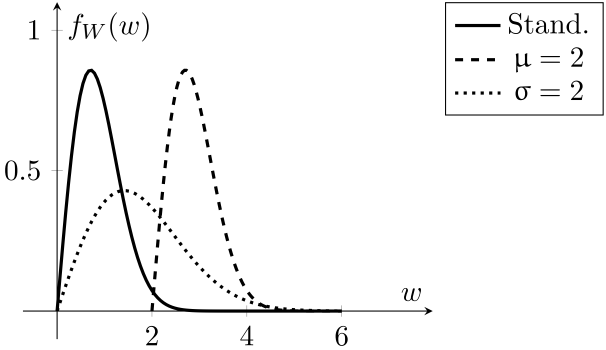
$$Q_W\left(p; \alpha, \mu, \sigma\right) = \left(-\log\left(1-p\right)\right)^{\frac{1}{\alpha}}$$

# The Weibull (Type III GEV) distribution (2/4)




- *Reverse* Weibull: the support is $\mathbb{Y} = (+\infty, \mu]$.

# The Weibull (Type III GEV) distribution (3/4)




- *Traditional* Weibull: the support is $\mathbb{W} = [\mu, \infty)$.

# The Weibull (Type III GEV) distribution (4/4)

## Observation 17

If $X \sim \text{Exp}(1)$, $Y = \mu - \sigma \log(X)$, and $W = \mu + \sigma X^{\frac{1}{\alpha}}$, it is as follows.

$$Y \sim \text{Gumbel}(\mu, \sigma) \quad \& \quad W \sim \text{Weibull}(\alpha, \mu, \sigma)$$

## Observation 18

If $X \sim \text{Exp}\left(\sqrt{\alpha}\right)$ and $W \sim \text{Weibull}\left(\alpha, 0, \frac{1}{2}\right)$, it is as follows.

$$X = \sqrt{W} \quad \& \quad W = X^2$$

## Observation 19

If $Y \sim \text{Frechet}(\alpha, \mu_Y, \sigma)$, and $W = (Y - \mu_Y)^{-1} + \mu_W$, it is as follows.

$$W \sim \text{Weibull}\left(\alpha, \mu_W, \sigma^{-1}\right)$$

A frequent application of the (traditional) Weibull distribution is in *survival analysis* (the statistical study of waiting times).