Statistics 1 Unit 6: Labs

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Kurt Hornik

Outline

- Heavy Tails
- Composite distributions
- Copulas

Tails

Terminology is somewhat messy.

The CDF $F(x) = \mathbb{P}(X \leq x)$ gives probabilities in the lower (left) tail; the complementary CDF (a.k.a. the survival function) $S(x) = \bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ gives probabilities in the upper (right) tail.

Remember the `lower.tail` argument to the p functions.

Tails

Terminology is somewhat messy.

The CDF $F(x) = \mathbb{P}(X \leq x)$ gives probabilities in the lower (left) tail; the complementary CDF (a.k.a. the survival function) $S(x) = \bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ gives probabilities in the upper (right) tail.

Remember the `lower.tail` argument to the p functions.

The *tail behavior* is what happens for $|x| \to \infty$.

I.e., lower/left tail probability $P(X \leq x)$ when $x \to -\infty$, upper/right tail probability $P(X > x)$ when $x \to \infty$.

Tails

Terminology is somewhat messy.

The CDF $F(x) = \mathbb{P}(X \leq x)$ gives probabilities in the lower (left) tail; the complementary CDF (a.k.a. the survival function) $S(x) = \bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ gives probabilities in the upper (right) tail.

Remember the `lower.tail` argument to the p functions.

The *tail behavior* is what happens for $|x| \to \infty$.

I.e., lower/left tail probability $P(X \leq x)$ when $x \to -\infty$, upper/right tail probability $P(X > x)$ when $x \to \infty$.

The tails are heavy when the tail probabilities decay slower than exponentially fast.

Tails of the normal distribution

One can show that as $x \to \infty$,

$$\mathbb{P}(N(0,1) > x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \sim e^{-x^2/2}.$$

Tails of the normal distribution

One can show that as $x \to \infty$,

$$\mathbb{P}(N(0,1) > x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \sim e^{-x^2/2}.$$

This goes to zero very rapidly, much faster than exponentially fast, i.e, $e^{-c|x|}$ for some $c > 0$.

Tails of the normal distribution

One can show that as $x \to \infty$,

$$\mathbb{P}(N(0,1) > x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \sim e^{-x^2/2}.$$

This goes to zero very rapidly, much faster than exponentially fast, i.e, $e^{-c|x|}$ for some $c > 0$.

Thus, the normal distribution has very light tails.

E.g.,

```
R> pnorm(1 : 10, lower.tail = FALSE)

 [1] 1.586553e-01 2.275013e-02 1.349898e-03 3.167124e-05 2.866516e-07
 [6] 9.865876e-10 1.279813e-12 6.220961e-16 1.128588e-19 7.619853e-24
```

Tails of the normal distribution

We can easily illustrate that $1 - \Phi(x) \approx e^{-x^2/2}$:

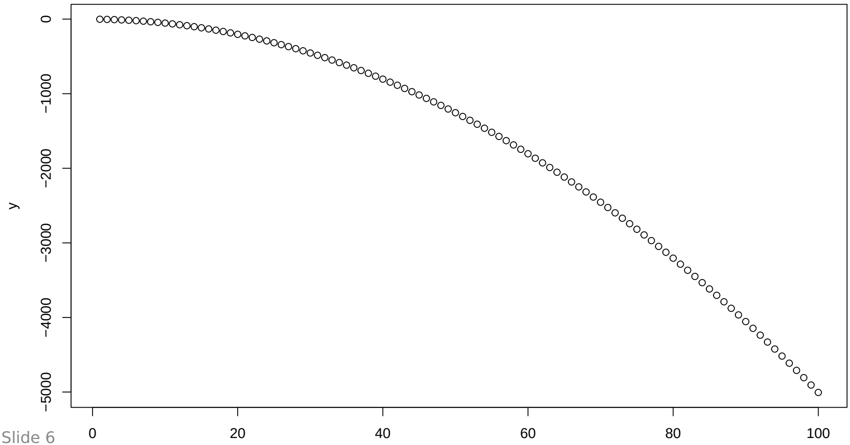
We take a sequence of x values and compute the corresponding $1 - \Phi(x)$ values, ideally on the log-scale for better precision:

```
R> x <- 1 : 100
R> y <- pnorm(x, lower.tail = FALSE, log.p = TRUE)
```

Tails of the normal distribution

Plot $\log(1 - \Phi(x))$ against x:

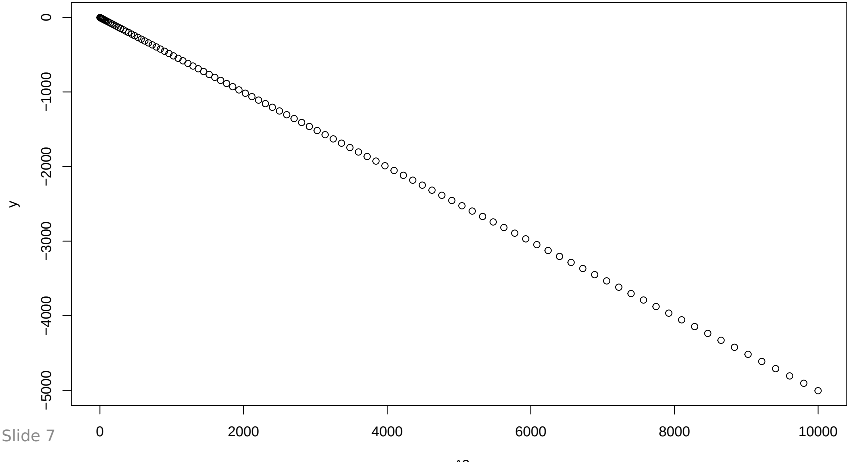
```
R> plot(x, y)
```



Tails of the normal distribution

Plot $\log(1 - \Phi(x))$ against x^2:

```
R> plot(x^2, y)
```



Tails of the exponential distribution

For the standard exponential distribution, trivially

$$\mathbb{P}(X > x) = \int_x^\infty e^{-t}\,dt = e^{-x}.$$

As $x \to \infty$, this goes to zero exponentially fast.

Tails of the exponential distribution

For the standard exponential distribution, trivially

$$\mathbb{P}(X > x) = \int_x^\infty e^{-t}\, dt = e^{-x}.$$

As $x \to \infty$, this goes to zero exponentially fast.

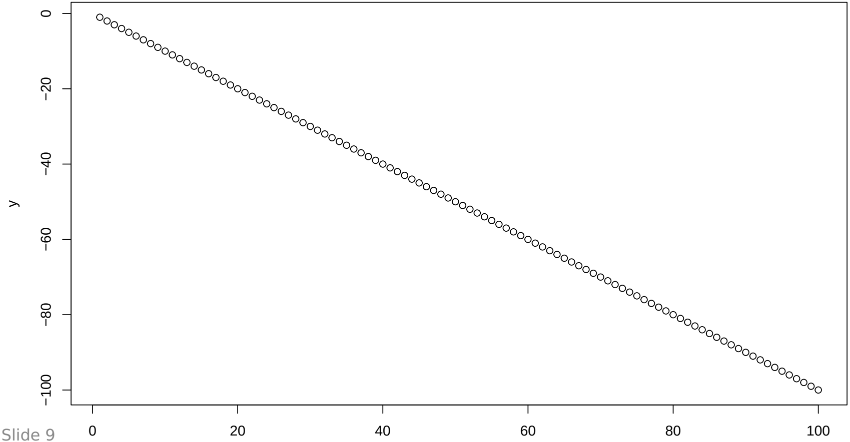
Again, we can illustrate using R as before:

```
R> x <- 1 : 100
R> y <- pexp(x, lower.tail = FALSE, log.p = TRUE)
```

Tails of the exponential distribution

Plot $\log(\mathbb{P}(X > x))$ against x:

```
R> plot(x, y)
```



Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Well, which continuous distributions do you remember?

Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Well, which continuous distributions do you remember? Normal,

Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Well, which continuous distributions do you remember? Normal, exponential,

Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Well, which continuous distributions do you remember? Normal, exponential, Gamma,

Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Well, which continuous distributions do you remember? Normal, exponential, Gamma, Uniform,

Heavy tails

Who knows a distribution with upper/right tail heavier than that of the exponential?

I.e., for which the upper/right tail probability does not go to zero exponentially fast?

Well, which continuous distributions do you remember? Normal, exponential, Gamma, Uniform, Student’s *t*!

Tails of the t distribution

For Student's t with n degrees of freedom, the density is

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

Tails of the t distribution

For Student's t with n degrees of freedom, the density is

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

As $x \to \infty$,

$$f(x) \sim (x^2)^{-(n+1)/2} = x^{-(n+1)} = \frac{1}{x^{n+1}}.$$

Hence,

$$1 - F(x) \approx \text{const} \int_x^\infty t^{-(n+1)}\, dt \sim x^{-n}.$$

Tails of the t distribution

This goes to zero *polynomially fast*, i.e., like a power of $1/x$.

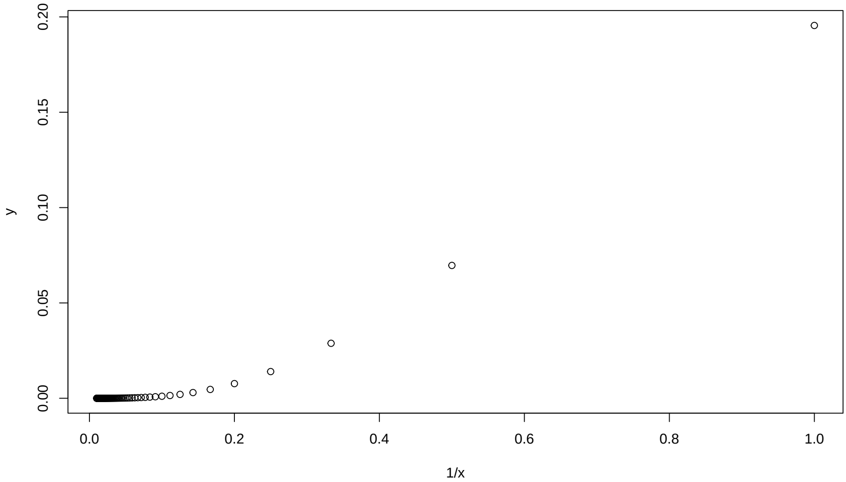
If we plot $1 - F(x)$ against $1/x$, things will look like a polynomial (for large x).

To illustrate for $n = 3$:

```
R> x <- 1 : 100
R> y <- pt(x, df = 3, lower.tail = FALSE)
```

Tails of the t distribution

```
R> plot(1 / x, y)
```



Tails of the t distribution

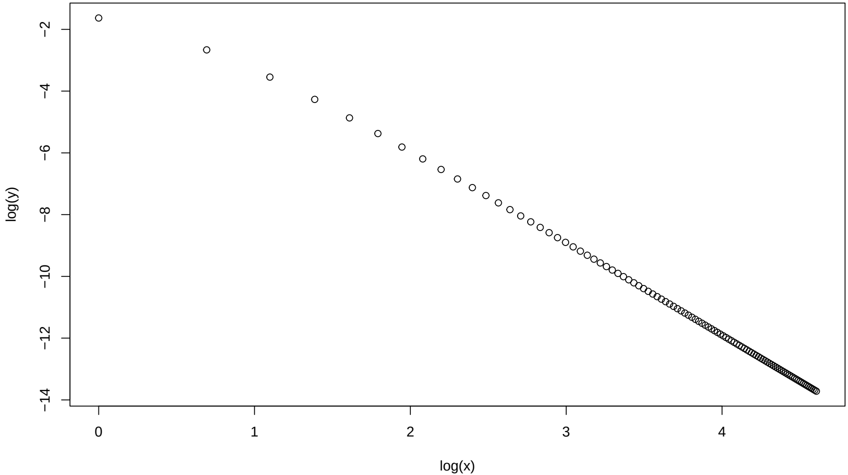
It is even better to take logs:

$$1 - F(x) \approx \text{const}\, x^{-n} \Rightarrow \log(1 - F(x)) \approx \log(\text{const}) - n \log(x)$$

So plotting $\log(1 - F(x))$ against $\log(x)$, things will look like a straight line with slope $-n$ (for large x).

Tails of the *t* distribution

```
R> plot(log(x), log(y))
```



Tails of the t distribution

Fitting a linear regression gives

```
R> lm(log(y) ~ log(x))

Call:
lm(formula = log(y) ~ log(x))

Coefficients:
(Intercept)       log(x)
    -0.3911      -2.8758
```

Tails of the t distribution

Fitting a linear regression gives

```
R> lm(log(y) ~ log(x))

Call:
lm(formula = log(y) ~ log(x))

Coefficients:
(Intercept)       log(x)
    -0.3911      -2.8758
```

We know that the slope should be -3: things would get better when using larger x.

Tails of the t distribution

We can also illustrate things by simulation (this will *not* work well for the normal or exponential distributions, as for these one needs very large n to draw large x values.

Tails of the t distribution

We can also illustrate things by simulation (this will *not* work well for the normal or exponential distributions, as for these one needs very large n to draw large x values.

Suppose we draw a sample of size n: $x_1, \dots, x_n$. Clearly,

$$\#\{x \text{ values} \geq k\text{-th largest } x \text{ value}\} = k.$$

So

$$1 - \hat{F}(k\text{-th largest value}) \approx k/n$$

and we can try to approximate the upper tail behavior by plotting $k/n, \dots, 1/n$ against the (suitably transformed) k largest x values for suitably chosen k.

Tails of the t distribution

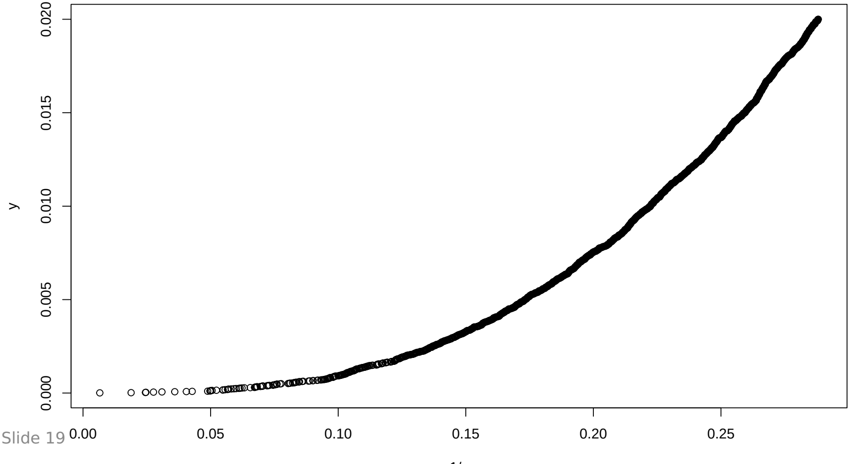
E.g., using $n = 100000$ and the upper 2% of the sample:

```
R> n <- 100000
R> z <- rt(n, df = 3)
R> z <- head(sort(z, decreasing = TRUE), 0.02 * n)
R> y <- seq_along(z) / n
```

Tails of the t distribution

Plot the empirical tail probabilities against the reciprocals largest values:

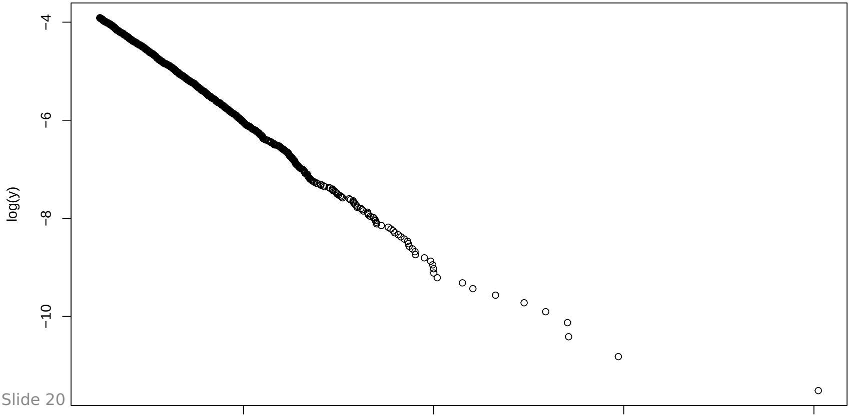
```
R> plot(1 / z, y)
```



Tails of the *t* distribution

Plot the logs of the empirical tail probabilities against the logs of the largest values:

```
R> plot(log(z), log(y))
```



Tails of the t distribution

A linear regression for the latter gives

```
R> lm(log(y) ~ log(z))

Call:
lm(formula = log(y) ~ log(z))

Coefficients:
(Intercept)       log(z)
    -0.4379      -2.7829
```

Not too bad.

Heavy-tailed distributions for QFin

We already saw that the tails of the weekly S&P 500 log-returns are much heavier than that of the normal distribution.

We can also see that they are much heavier than that of the exponential distribution.

This is actually a very important observation!

Why?

Heavy-tailed distributions for QFin

- Mean-variance portfolio theory works best for normally distributed returns (issue for asset management)

Heavy-tailed distributions for QFin

- Mean-variance portfolio theory works best for normally distributed returns (issue for asset management)
- In risk management, we need to estimate small/large quantiles (Value at Risk) and the corresponding conditional expectations (Expected Shortfall). Assuming normality, we would under-estimate these (Basel) risk parameters.

Heavy-tailed distributions for QFin

- Mean-variance portfolio theory works best for normally distributed returns (issue for asset management)
- In risk management, we need to estimate small/large quantiles (Value at Risk) and the corresponding conditional expectations (Expected Shortfall). Assuming normality, we would under-estimate these (Basel) risk parameters.
- However, direct estimation may not work well, as empirical estimation of small/large quantiles can only use few data and hence has low precision.

Heavy-tailed distributions for QFin

- Mean-variance portfolio theory works best for normally distributed returns (issue for asset management)
- In risk management, we need to estimate small/large quantiles (Value at Risk) and the corresponding conditional expectations (Expected Shortfall). Assuming normality, we would under-estimate these (Basel) risk parameters.
- However, direct estimation may not work well, as empirical estimation of small/large quantiles can only use few data and hence has low precision.
- Suggests using suitable parametric models for the tails.

Heavy-tailed distributions for QFin

To appropriately model heavy-tailed return distributions:

- What are good parametric models for heavy tails?

Heavy-tailed distributions for QFin

To appropriately model heavy-tailed return distributions:

- What are good parametric models for heavy tails?
  Student's t is one possibility, but there is a problem: it is *symmetric*, and usually the loss and profit tails are not.

Heavy-tailed distributions for QFin

To appropriately model heavy-tailed return distributions:

- What are good parametric models for heavy tails?
  Student's t is one possibility, but there is a problem: it is *symmetric*, and usually the loss and profit tails are not.
  There are asymmetric generalizations, but there are also simpler alternatives: the Pareto and the Generalized Pareto distributions.

Heavy-tailed distributions for QFin

To appropriately model heavy-tailed return distributions:

- What are good parametric models for heavy tails?
  Student's *t* is one possibility, but there is a problem: it is *symmetric*, and usually the loss and profit tails are not.
  There are asymmetric generalizations, but there are also simpler alternatives: the Pareto and the Generalized Pareto distributions.
- The normal works well for the center, but not for the tails: how can we piece things together from separate models for the center and the left/right tails?

Heavy-tailed distributions for QFin

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

To appropriately model heavy-tailed return distributions:

- What are good parametric models for heavy tails?
  Student's t is one possibility, but there is a problem: it is *symmetric*, and usually the loss and profit tails are not.
  There are asymmetric generalizations, but there are also simpler alternatives: the Pareto and the Generalized Pareto distributions.
- The normal works well for the center, but not for the tails: how can we piece things together from separate models for the center and the left/right tails?
  This needs so-called *composite distributions*.

Pareto distributions

The Pareto with *scale* parameter $K > 0$ and *shape* parameter $\alpha > 0$ has CDF

$$F(x) = \begin{cases} 1 - (K/x)^{\alpha}, & x \geq K, \\ 0, & x < K. \end{cases}$$

Hence, for $x \geq K$,

$$1 - F(x) = (K/x)^{\alpha}$$

(so exactly proportional to $x^{-\alpha}$).

Pareto distributions

The Pareto with *scale* parameter $K > 0$ and *shape* parameter $\alpha > 0$ has CDF

$$F(x) = \begin{cases} 1-(K/x)^{\alpha}, & x \geq K, \\ 0, & x < K. \end{cases}$$

Hence, for $x \geq K$,

$$1-F(x) = (K/x)^{\alpha}$$

(so exactly proportional to $x^{-\alpha}$).

Note that K clearly is a scale parameter, but also gives the Kut-off where the support starts, so in some sense the location. Strange.

Generalized Pareto distributions

The generalized Pareto distribution (GPD) with location parameter μ, scale parameter $\sigma > 0$ and shape parameter ξ has CDF

$$F(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x-\mu}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{x-\mu}{\sigma}\right), & \xi = 0, \end{cases}$$

for $x \geq \mu$ if $\xi \geq 0$ and $\mu \leq x \leq \mu - \sigma/\xi$ if $\xi < 0$.

Generalized Pareto distributions

The generalized Pareto distribution (GPD) with location parameter μ, scale parameter $\sigma > 0$ and shape parameter ξ has CDF

$$F(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x-\mu}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{x-\mu}{\sigma}\right), & \xi = 0, \end{cases}$$

for $x \geq \mu$ if $\xi \geq 0$ and $\mu \leq x \leq \mu - \sigma/\xi$ if $\xi < 0$.

See the homeworks for a bit more on this.

Generalized Pareto distributions

The generalized Pareto distribution (GPD) with location parameter μ, scale parameter $\sigma > 0$ and shape parameter ξ has CDF

$$F(x) = \begin{cases} 1 - \left(1 + \xi \dfrac{x-\mu}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\dfrac{x-\mu}{\sigma}\right), & \xi = 0, \end{cases}$$

for $x \geq \mu$ if $\xi \geq 0$ and $\mu \leq x \leq \mu - \sigma/\xi$ if $\xi < 0$.

See the homeworks for a bit more on this.

Clearly, μ is a location parameter, and σ is a scale parameter.

Generalized Pareto distributions

For $\xi > 0$, as $x \to \infty$

$$1 - F(x) = \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \sim x^{-1/\xi}$$

so $1/\xi$ corresponds to α (for the Pareto) or n (for Student's t).

Generalized Pareto distributions

For $\xi > 0$, as $x \to \infty$

$$1 - F(x) = \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \sim x^{-1/\xi}$$

so $1/\xi$ corresponds to α (for the Pareto) or n (for Student's t).

The larger ξ, the heavier the tails.

Generalized Pareto distributions

For $\xi > 0$, as $x \to \infty$

$$1 - F(x) = \left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi} \sim x^{-1/\xi}$$

so $1/\xi$ corresponds to α (for the Pareto) or n (for Student's t).

The larger ξ, the heavier the tails.

Remember that as $n \to \infty$,

$$\left(1 + \frac{z}{n}\right)^n \to e^z.$$

Replacing n by $1/\xi$, as $\xi \to 0+$ (the plus means "from the right"),

$$(1 + \xi z)^{1/\xi} \to e^z$$

Generalized Pareto distributions

Hence, with $z = (x - \mu)/\sigma$, as $\xi \to 0+$,

$$1 - F(x) = (1 + \xi z)^{-1/\xi} \to e^{-z} = e^{-(x-\mu)/\sigma}.$$

This explains the definition of the GPD for $\xi = 0$.

For $\xi = 0$, we get the location-scale family generated by the standard exponential distribution. This (trivially) has light tails.

For connaiseurs: for $\xi < 0$, the GPD has compact support. Thus, very light tails.

Outline

- Heavy Tails
- Composite distributions
- Copulas

Composite distributions

If X has distribution function F and $B = (\alpha, \beta]$ is an interval, then for $\alpha < x \leq \beta$,

$$\mathbb{P}(X \leq x | x \in B) = \frac{\mathbb{P}(\alpha < X \leq x)}{\mathbb{P}(\alpha < X \leq \beta)} = \frac{F(x) - F(\alpha)}{F(\beta) - F(\alpha)}.$$

This is the conditional distribution function $F(x|B)$ of X given $X \in B$.

Composite distributions

If X has distribution function F and $B = (\alpha, \beta]$ is an interval, then for $\alpha < x \leq \beta$,

$$\mathbb{P}(X \leq x | x \in B) = \frac{\mathbb{P}(\alpha < X \leq x)}{\mathbb{P}(\alpha < X \leq \beta)} = \frac{F(x) - F(\alpha)}{F(\beta) - F(\alpha)}.$$

This is the conditional distribution function $F(x|B)$ of X given $X \in B$.

If X has density f, then this conditional distribution has density

$$f(x|B) = \frac{f(x)}{F(\beta) - F(\alpha)}, \quad x \in B.$$

Composite distributions

If $B_1, \ldots, B_k$ is a partition of $\mathbb{R}$ into disjoint intervals,

$$\mathbb{P}(X \leq x) = \sum_i \mathbb{P}(X \in B_i)\mathbb{P}(X \leq x | X \in B_i) = \sum w_i F(x|B_i)$$

Composite distributions

If $B_1, \ldots, B_k$ is a partition of $\mathbb{R}$ into disjoint intervals,

$$\mathbb{P}(X \leq x) = \sum_i \mathbb{P}(X \in B_i)\mathbb{P}(X \leq x | X \in B_i) = \sum w_i F(x|B_i)$$

This expresses F as the mixture of the conditional distributions.

This is a special mixture where the components have disjoint support.
These mixtures are called *composite distributions*.

Composite distributions

If $B_1, \ldots, B_k$ is a partition of $\mathbb{R}$ into disjoint intervals,

$$\mathbb{P}(X \leq x) = \sum_i \mathbb{P}(X \in B_i)\mathbb{P}(X \leq x | X \in B_i) = \sum w_i F(x|B_i)$$

This expresses F as the mixture of the conditional distributions.

This is a special mixture where the components have disjoint support. These mixtures are called *composite distributions*.

If X has density f, then

$$f(x) = \sum_i w_i f(x|B_i).$$

Composite distributions

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

For each i, the component distribution (or density) can be obtained as an arbitrary distribution (or density) conditioned to be in B_i.

I.e., if we have distributions $G_1, \dots, G_k$ and weights $w_1, \dots, w_k$, then

$$F(x) = \sum_i w_i G_i(x|B_i)$$

is a composite distribution.

If $X \sim F$ and $Y_1 \sim G_1, \dots, Y_k \sim G_k$, then the distribution of X conditional on $X \in B_i$ is the same as the distrubution of Y_i conditional on $Y_i \in B_i$.

(A complicated way of interpreting the above equation.)

Composite distributions

For each i, the component distribution (or density) can be obtained as an arbitrary distribution (or density) conditioned to be in B_i.

I.e., if we have distributions $G_1, \dots, G_k$ and weights $w_1, \dots, w_k$, then

$$F(x) = \sum_i w_i G_i(x|B_i)$$

is a composite distribution.

If $X \sim F$ and $Y_1 \sim G_1, \dots, Y_k \sim G_k$, then the distribution of X conditional on $X \in B_i$ is the same as the distrubution of Y_i conditional on $Y_i \in B_i$.

(A complicated way of interpreting the above equation.)

Similarly for densities,

$$f(x) = \sum_i w_i g_i(x|B_i).$$

Composite distributions

In R, composites (and more general mixtures) can be done very conveniently using package `mistr`.

This has functions for generating general composites and fitting the models we need for heavy-tailed financial returns, and then `d`, `p`, `q` and `r` functions for composites (and mixtures).

Composite distributions

To model our weekly S&P 500 log-returns, we need something heavy-tailed in the left and right tail, and can use the normal in the center.

One way of doing is taking a composite of three distributions.

I.e., $k = 3$, and

$$B_1 = (-\infty, \beta_1], \quad B_2 = (\beta_1, \beta_2], \quad B_3 = (\beta_2, \infty)$$

for suitable break points β_1 and β_2.

Composite distributions

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

To model our weekly S&P 500 log-returns, we need something heavy-tailed in the left and right tail, and can use the normal in the center.

One way of doing is taking a composite of three distributions.

I.e., $k = 3$, and

$$B_1 = (-\infty, \beta_1], \quad B_2 = (\beta_1, \beta_2], \quad B_3 = (\beta_2, \infty)$$

for suitable break points β_1 and β_2. If the component distributions have densities g_1, g_2 and g_3, then the density of the composite is

$$f(x) = \begin{cases} w_1 \frac{g_1(x)}{G_1(\beta_1)}, & x \le \beta_1, \\ w_2 \frac{g_2(x)}{G_2(\beta_2) - G_2(\beta_1)}, & \beta_1 < x \le \beta_2, \\ w_3 \frac{g_3(x)}{1 - G_3(\beta_2)}, & x > \beta_2. \end{cases}$$

Composite distributions

Clearly,

$$F(\beta_1) = w_1, \quad F(\beta_2) = w_1 + w_2 \Rightarrow w_2 = F(\beta_2) - F(\beta_1).$$

When fitting such models to data $x_1, \ldots, x_n$, naturally we take the w_i to match the empirical frequencies in the intervals, i.e.,

$$\begin{aligned}
\hat{w}_1 &= \hat{F}_n(\beta_1) &= \frac{\#\{i : x_i \le \beta_1\}}{n}, \\
\hat{w}_2 &= \hat{F}_n(\beta_2) - \hat{F}_n(\beta_1) &= \frac{\#\{i : \beta_1 < x_i \le \beta_2\}}{n}, \\
\hat{w}_3 &= 1 - \hat{w}_1 - \hat{w}_2 &= \frac{\#\{i : x_i > \beta_2\}}{n}.
\end{aligned}$$

The PNP model

The simplest such model is taking Pareto distributions for the tails and the normal for the center.

The PNP model

The simplest such model is taking Pareto distributions for the tails and the normal for the center. More precisely,

$$\begin{aligned} G_1 &= \text{negative of Pareto}(K_1, \alpha_1), \\ G_2 &= \text{Normal}(\mu, \sigma), \\ G_3 &= \text{Pareto}(K_2, \alpha_2). \end{aligned}$$

The PNP model

The simplest such model is taking Pareto distributions for the tails and the normal for the center. More precisely,

$$\begin{aligned} G_1 &= \text{negative of Pareto}(K_1, \alpha_1), \\ G_2 &= \text{Normal}(\mu, \sigma), \\ G_3 &= \text{Pareto}(K_2, \alpha_2). \end{aligned}$$

This seems to have parameters

$$K_1, \alpha_1, \quad \mu, \sigma, \quad K_2, \alpha_2 \qquad \text{and the breakpoints } \beta_1, \beta_2.$$

Naturally (but see the notes), $K_1 = -\beta_1$ and $K_2 = \beta_2$. If we require the density to be continuous at β_1 and β_2, we get 2 more restrictions, for 4 free parameters. In `mistr`: `PNP_fit()`.

The GNG model

Alternatively, we could take GPDs for the tails and a normal for the center.

The GNG model

Alternatively, we could take GPDs for the tails and a normal for the center. More precisely,

$$
\begin{aligned}
G_1 &= \text{negative of GPD}(\mu_1, \sigma_1, \xi_1), \\
G_2 &= \text{Normal}(\mu, \sigma), \\
G_3 &= \text{GPD}(\mu_2, \sigma_2, \xi_2).
\end{aligned}
$$

The GNG model

Alternatively, we could take GPDs for the tails and a normal for the center. More precisely,

$$
\begin{aligned}
G_1 &= \text{negative of GPD}(\mu_1, \sigma_1, \xi_1), \\
G_2 &= \text{Normal}(\mu, \sigma), \\
G_3 &= \text{GPD}(\mu_2, \sigma_2, \xi_2).
\end{aligned}
$$

This seems to have parameters

$$\mu_1, \sigma_1, \xi_1, \quad \mu, \sigma, \quad \mu_2, \sigma_2, \xi_2 \qquad \text{and the breakpoints } \beta_1, \beta_2.$$

Again, naturally $\mu_1 = -\beta_1$ and $\mu_2 = \beta_2$, Again, requiring continuity adds 2 more restrictions, for 6 free parameters. In `mistr`: `GNG_fit()`.

Outline

- Heavy Tails
- Composite distributions
- Copulas

Quantiles and PITs

We already know:

Let F be a distribution function and Q_F be its quantile function.

If $U \sim U_{0,1}$, then for the quantile transform $X = Q_F(U)$ we have

$$X = Q_F(U) \sim F.$$

(This we proved in class.)

If $X \sim F$ and F is continuous, then for the probability integral transform (PIT) $U = F(X)$ we have

$$U = F(X) \sim U_{0,1}.$$

(This we proved in the homeworks.)

Copulas

Now consider a pair (X, Y) of random variables.

Write F_X and F_Y for their marginal distributions and G for their joint distribution. I.e.,

$$F_X(x) = \mathbb{P}(X \le x), \quad F_Y(y) = \mathbb{P}(Y \le y), \quad G(x, y) = \mathbb{P}(X \le x, Y \le y).$$

Copulas

Now consider a pair (X, Y) of random variables.

Write F_X and F_Y for their marginal distributions and G for their joint distribution. I.e.,

$$F_X(x) = \mathbb{P}(X \leq x), \quad F_Y(y) = \mathbb{P}(Y \leq y), \quad G(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

If F_X and F_Y are continuous, we know for the PITs that

$$U = F_X(X) \sim U_{0,1}, \quad V = F_Y(Y) \sim U_{0,1}.$$

Copulas

Now consider a pair (X, Y) of random variables.

Write F_X and F_Y for their marginal distributions and G for their joint distribution. I.e.,

$$F_X(x) = \mathbb{P}(X \leq x), \quad F_Y(y) = \mathbb{P}(Y \leq y), \quad G(x, y) = \mathbb{P}(X \leq x, Y \leq y).$$

If F_X and F_Y are continuous, we know for the PITs that

$$U = F_X(X) \sim U_{0,1}, \quad V = F_Y(Y) \sim U_{0,1}.$$

We can ask: what is the joint distribution of U and V? This is the *copula* of X and Y (or G).

Copulas

A (bivariate) copula is the joint CDF of a pair (U, V) with standard uniform margins.

(This obviously generalizes to d-variate.)

I.e., writing C for the copula,

$$C(u, v) = \mathbb{P}(U \leq u, V \leq v).$$

Hence, for $0 \leq u, v \leq 1$,

$$C(u, 0) = 0, \quad C(0, v) = 0, \qquad C(u, 1) = u, \quad C(1, v) = v.$$

In particular, $C(1, 1) = 1$.

Copulas

The simplest copulas are the ones where

- $V = U$ (comonotonicity copula)

Copulas

The simplest copulas are the ones where

- $V = U$ (comonotonicity copula)
- $V = 1 - U$ (countermonotonicity copula)

Copulas

The simplest copulas are the ones where

- $V = U$ (comonotonicity copula)
- $V = 1 - U$ (countermonotonicity copula)
- U and V are independent (independence copula).

If $V = U$ (perfect positive dependence),

$$\mathbb{P}(U \leq u, V \leq v) = \mathbb{P}(U \leq u, U \leq v) = \mathbb{P}(U \leq \min(u, v)) = \min(u, v)$$

so the comonotonicity copula is

$$C(u, v) = \min(u, v), \quad 0 \leq u, v, \leq 1.$$

Copulas

If $V = 1 - U$ (perfect negative dependence),

$$\begin{aligned}
\mathbb{P}(U \leq u, V \leq v) &= \mathbb{P}(U \leq u, 1 - U \leq v) \\
&= \mathbb{P}(1 - v \leq U \leq u) \\
&= \begin{cases} u - (1 - v) & \text{if } u - (1 - v) \geq 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \max(u + v - 1, 0)
\end{aligned}$$

so the countermonotonicity copula is

$$C(u, v) = \max(u + v - 1, 0), \quad 0 \leq u, v, \leq 1.$$

Copulas

Finally, if U and V are independent:

$$P(U \leq u, V \leq v) = \mathbb{P}(U \leq u)\mathbb{P}(V \leq v) = uv,$$

so the independence copula is

$$C(u, v) = uv, \quad 0 \leq u, v, \leq 1.$$

Copulas

All bivariate distributions with continuous margins have corresponding copulas.

E.g., the bivariate normal distribution has 5 parameters: $\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho_{XY}$.

The first 4 relate to the margins, so the bivariate normal (or Gauss) copula has one parameter: ρ_{XY}:

$$C(u,v) = \mathbb{P}\left(\Phi\left(\frac{X-\mu_X}{\sigma_X}\right) \leq u, \Phi\left(\frac{Y-\mu_Y}{\sigma_Y}\right) \leq v\right).$$

Copulas

Which copula to use for modeling dependence?

Copulas

Which copula to use for modeling dependence? It depends . . .

For risk management we usually need something with (loss) *tail dependence* (if large losses co-occur more often than under independence).

Copulas

Which copula to use for modeling dependence? It depends . . .

For risk management we usually need something with (loss) *tail dependence* (if large losses co-occur more often than under independence).

For our Coffee data set, we use package `VineCopula` which conveniently selects the "best" copula from a set of candidates.

Copulas

Sklar's theorem says: for all joint distributions G there is a copula C such that

$$G(x, y) = C(F_X(x), F_Y(y)).$$

In the continuous case, C is unique.

In the d-variate case, with

$$G(x_1, \ldots, x_d) = \mathbb{P}(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

and

$$F_i(x_i) = \mathbb{P}(X_i \leq x_i), \quad 1 \leq i \leq d,$$

we can always do

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

Copulas

Suppose we can draw (U, V) from the copula of X and Y.

We can then draw X and Y via the quantile transforms of U and V.

I.e.:

$$(U, V) \sim C, \qquad X = Q_{F_X}(U), Y = Q_{F_Y}(V)$$

Copulas

Suppose we can draw (U, V) from the copula of X and Y.

We can then draw X and Y via the quantile transforms of U and V.

I.e.:

$$(U, V) \sim C, \qquad X = Q_{F_X}(U), Y = Q_{F_Y}(V)$$

To make sure:

$$\begin{aligned} \mathbb{P}(Q_{F_X}(U) \leq x, Q_{F_Y}(V) \leq y) &= \mathbb{P}(U \leq F_X(x), V \leq F_Y(y)) \\ &= C(F_X(x), F_Y(y)) \\ &= G(x, y). \end{aligned}$$

Copulas

Clearly, if F_X and F_Y are invertible:

$$Q_{F_X}(u) = x \Leftrightarrow u = F_X(x), \qquad Q_{F_Y}(v) = y \Leftrightarrow v = F_Y(y).$$

For the copula, we have

$$C(u, v) = \mathbb{P}(F_X(X) \le u, F_Y(Y) \le v) = \mathbb{P}(X \le Q_{F_X}(u), Y \le Q_{F_Y}(v))$$

and thus

$$\mathbb{P}(X \le x, Y \le y) = C(F_X(x), F_Y(y)).$$

(Sklar's theorem says that C is unique in this case, and in fact whenever things are continuous.)