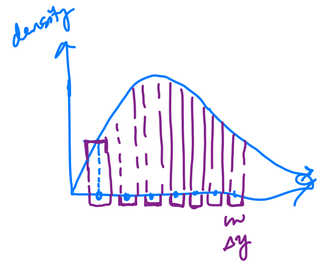
To get a probability, need to go from a density to a measure. i.e. integrate!

But likelihood not necessarily going to integrate to 1 (won't necessarily return value in [0,1]).

But it may be useful to think of the value of the likelihood as being proportional to a probability.

Given data $y_1, \ldots, y_n$

$$L(\alpha, \beta \mid y) = \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} y_i^{\alpha-1} e^{-y_i/\beta}$$



for α, β given

prob is approximately $\prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha} y_i^{\alpha-1} e^{-y_i/\beta} \Delta y$.

More formally, begin with the definition of a derivative

$$g'(x) = \lim_{h \to 0^+} \frac{g(x+h) - g(x-h)}{2h}.$$

Let F be the cumulative distribution function of a continuous random variable Y, then (if the derivative exists)

$$f(y) = \lim_{h \to 0^+} \frac{F(y+h) - F(y-h)}{2h} = \lim_{h \to 0^+} \frac{P(Y \in (y-h, y+h))}{2h}$$

If we substitute this definition of a density into the definition of the likelihood

$$L(\underline{\theta} \mid \underline{Y}) = \prod_{i=1}^{n} f(Y_i; \underline{\theta})$$

$$= \prod_{i=1}^{n} \lim_{h \to 0^+} \frac{F(Y_i + h; \theta) - F(Y_i - h; \theta)}{2h}$$

$$= \lim_{h \to 0^+} \prod_{i=1}^{n} \frac{1}{2h} \left(F(Y_i + h; \theta) - F(Y_i - h; \theta) \right)$$

$$= \lim_{h \to 0^+} \underbrace{\left(\frac{1}{2h}\right)^n}_{} \overbrace{\prod_{i=1}^{n} P_{\underline{\theta}} \left(Y_i^* \in (Y_i - h, Y_i + h] \right)}^{}$$

likelihood is proportional to a probability.

as $h \to 0^+$, probability $\downarrow 0$ and $\left(\frac{1}{2h}\right)^n \uparrow$ to balance.

proportionality "constant" is constant b/c doesn't depend on $\underline{\theta}$.

Y_i^* independent of Y_i's w/ same dsn as Y_i's.

$\Rightarrow$ likelihood is proportional to the probability of obtaining a new sample that is close to sample obtained.

Compare this to the iid discrete case:

$$L(\underset{\sim}{\theta}|\underset{\sim}{y}) = \prod_{i=1}^{n} f(y_i;\underset{\sim}{\theta})$$

$$= \lim_{h\to 0+} \prod_{i=1}^{n} \{F(y_i+h;\underset{\sim}{\theta}) - F(y_i-h;\underset{\sim}{\theta})\}$$

now we don't need the proportionality constant.

So likelihoods for discrete RVs get weighted differently than likelihoods for continuous RV's.

This has to do w/ underlying dominating measure of these RV's.
(counting measure & Lebesgue measure respectively).

Thought for later: What about mixtures?

Example (Hurricane Data, Cont'd): Recall with a gamma model, the likelihood for this example is

$$L(\boldsymbol{\theta}|\boldsymbol{Y}) = \{\Gamma(\alpha)\}^{-n}\beta^{-n\alpha}\left\{\prod Y_i\right\}^{\alpha-1}\exp\left(-\sum y_i/\beta\right),$$

how to take derivatives, set $=0$, solve?

and log-likelihood

$$\ell(\boldsymbol{\theta}) = -n\log\Gamma(\alpha) - n\alpha\log\beta + (\alpha-1)\sum\log Y_i - \sum Y_i/\beta$$

$$\frac{\partial \log\Gamma(\alpha)}{\partial\alpha} = -\gamma + \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{k+\alpha-1}\right)$$

Euler-Mascheroni constant.

```
## loglikelihood function
neg_gamma_loglik <- function(theta, data) {
  -sum(log(dgamma(data, theta[1], scale = theta[2])))
}

## maximize
mle <- nlm(neg_gamma_loglik, c(1.59, 4.458), data =
hurr_rain)
mle$estimate
```

partial derivatives of Gamma likelihood do not result in a linear system of equations $\Rightarrow$ use numerical optimization.

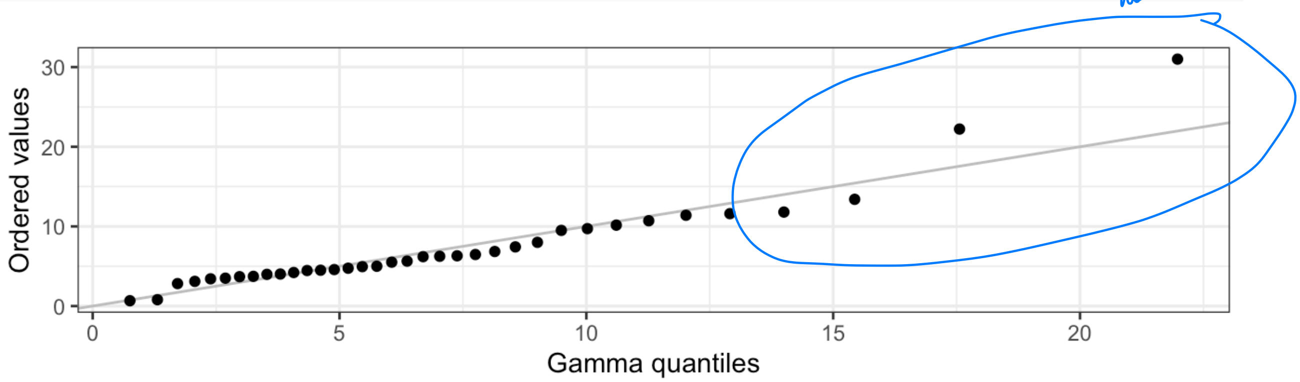
or "optim"

$\hat{\alpha}_{MLE}$ $\hat{\beta}_{MLE}$

```
## [1] 2.187214 3.331862
```

```
## Gamma QQ plot
data.frame(theoretical = qgamma(ppoints(hurr_rain),
mle$estimate[1], scale = mle$estimate[2]),
           actual = sort(hurr_rain)) |>
  ggplot() +
  geom_abline(aes(intercept = 0, slope = 1), colour =
"grey") +
  geom_point(aes(theoretical, actual)) +
  xlab("Gamma quantiles") + ylab("Ordered values")
```

not a great fit here.



1.2.4 Mixtures of Discrete and Continuous RVs

Some data Y often have a number of zeros and the amounts greater than zero are best modeled by a continuous distribution.

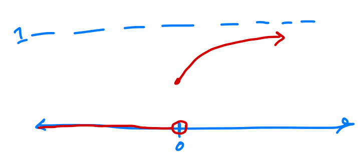
Ex: Rainfall, Snowfall, Amount of time I climb in a day, etc.

In other words, they have positive probability of taking a value of exactly zero, but continuous distribution otherwise.

- different than ZIP, which is discrete.
- can be generalized to other cases, point mass doesn't have to be at zero (e.g. record linkage models)

A sensible model would assume Y_i are iid with cdf

$$F_Y(y; p, \boldsymbol{\theta}) = \begin{cases} 0 & y < 0 \\ p & y = 0 \\ p + (1-p)F_T(y; \boldsymbol{\theta}) & y > 0 \end{cases}$$



where $0 < p \leq 1$ is $P(Y = 0)$ and $F_T(y; \boldsymbol{\theta})$ is a distribution function for a continuous positive random variable.

Another way to write this:

$$F_Y(y; p, \underline{\theta}) = P(Y \leq y) = p\,\mathbb{I}(0 \leq y) + (1-p)F_T(y; \underline{\theta}).$$

No problems w/ cdf.

How to go from here to get a likelihood?

(what is the "density" here?)

practical.

One approach: let n_0 be the number of zeroes in the data and $m = n - n_0$ be the number of non-zero Y_i. This leads to an intuitive way to construct the likelihood for iid $Y_1, \ldots, Y_n$ distributed according to the above distribution:

combining the constructions for continuous/discrete RVs.

$$L(\overset{p,}{\boldsymbol{\theta}}|\boldsymbol{Y}) = \lim_{h\to 0^+} \overbrace{\left(\frac{1}{2h}\right)^m}^{\text{weights for continuous part}} \prod_{i=1}^{n} \{F_Y(Y_i + h; p, \boldsymbol{\theta}) - F_Y(Y_i - h; p, \boldsymbol{\theta})\}$$

$$= \lim_{h\to0^+} \{F_Y(h; p, \underline{\theta}) - \underbrace{F_Y(-h; p, \underline{\theta})}_{=0}\}^{n_0} \times \lim_{h\to0^+} \prod_{Y_i>0} \left\{\frac{F_Y(Y_i+h; p, \underline{\theta}) - F_Y(Y_i-h; p, \underline{\theta})}{2h}\right\}$$

for $Y_i=0$; m Y_i's that >0.

$$= \lim_{h\to0^+} \{p + (1-p)F_T(h; \underline{\theta})\}^{n_0} \times \lim_{h\to0^+} \prod_{Y_i>0} \left\{\frac{p + (1-p)F_T(Y_i+h; \underline{\theta}) - p - (1-p)F_T(Y_i-h; p, \underline{\theta})}{2h}\right\}$$

$$= \underbrace{p^{n_0} \times (1-p)^m}_{\substack{\text{Bernoulli component} \\ \text{for } n_0 = n-m \text{ zeroes.}}} \underbrace{\prod_{Y_i>0} f_T(Y_i; \underline{\theta})}_{\substack{\text{continuous component for} \\ m \text{ non-zero values}}}$$

$$\ell(p, \underline{\theta}) = (n-m)\log p + m\log(1-p) + \sum_{Y_i>0} \log f_T(Y_i; \underline{\theta})$$

Notice: $\hat{p}_{MLE} = \frac{n-m}{n}$ & $\hat{\underline{\theta}}_{MLE}$ obtained in usual way from m obs w/ density $f_T(y; \underline{\theta})$.

will only use $Y_i > 0$

Kind of like law of total probability (informally):

$$L(p, \underline{\theta}|\underline{Y}) \propto \prod_{i=1}^{n} P(Y_i = y)$$

$$= \prod_{i=1}^{n} \{P(Y_i = y_i | Y_i = 0)P(Y_i = 0) + P(Y_i = y_i | Y_i \neq 0)P(Y_i \neq 0)\}$$

$$= \prod_{i=1}^{n} \{\delta_{Y_i=0}(y_i)\, p + \delta_{Y_i\neq0}(y_i)\, f_T(y_i|\underline{\theta}) \cdot (1-p)\}$$

$$= p^{n-m}(1-p)^m \prod_{Y_i>0} f_T(y_i|\underline{\theta}).$$

Feels a little arbitrary in how we are defining different weights on our likelihood for discrete and continuous parts.

Turns out, it doesn't matter! (Need some STAT 630/720 to see why.)

not much!

small aside:

Definition (Absolute Continuity) On $(\mathbb{X}, \mathcal{M})$, a finitely additive set function ϕ is *absolutely continuous* with respect to a measure μ if $\phi(A) = 0$ for each $A \in \mathcal{M}$ with $\mu(A) = 0$. We also say ϕ is *dominated* by μ and write $\phi \ll \mu$. If ν and μ are measures such that $\nu \ll \mu$ and $\mu \ll nu$ then μ and ν are *equivalent*.

ν

Ex: a discrete distribution is dominated by the counting measure

a continuous distribution is dominated by the Lebesgue measure.

Theorem (Lebesgue-Randon-Nikodym) Assume that ϕ is a σ-finite countably additive set function and μ is a σ-finite measure. There exist unique σ-finite countably additive set functions ϕ_s and ϕ_{ac} such that $\phi = \phi_{ac} + \phi_s$, $\phi_{ac} \ll \mu$, ϕ_s and μ are mutually singular and there exists a measurable extended real valued function f such that

$\exists$ a set B s.t. $\phi_s(B) = 0$ and $\mu(B^c) = 0$.

$$\phi_{ac}(A) = \int_A f d\mu, \qquad \text{for all } A \in \mathcal{M}.$$

If g is another such function, then $f = g$ a.e. wrt μ. If $\phi \ll \mu$ then $\phi(A) = \int_A f d\mu$ for all $A \in \mathcal{M}$.

let μ = Lebesgue

ν = counting measure over $\{0\}$.

Think about a measure ϕ w/ positive value at 0 and also continuous ≥ 0.

$\Rightarrow \mu(\{0\}) = 0$ but $\nu(\{0\}^c) = 0$ $\Rightarrow \phi_s = \nu$ and ϕ_{ac} is the rest.

Definition (Radon-Nikodym Derivative) $\phi = \phi_{ac} + \phi_s$ is called the *Lebesgue decomposition*. If $\phi \ll \mu$, then the density function f is called the *Radon-Nikodym derivative* of ϕ wrt μ.

"density"

So what?