# Pooling and Meta-analysis

Tony O'Hagan

# Pooling

Synthesising prior information from several experts

# Multiple experts

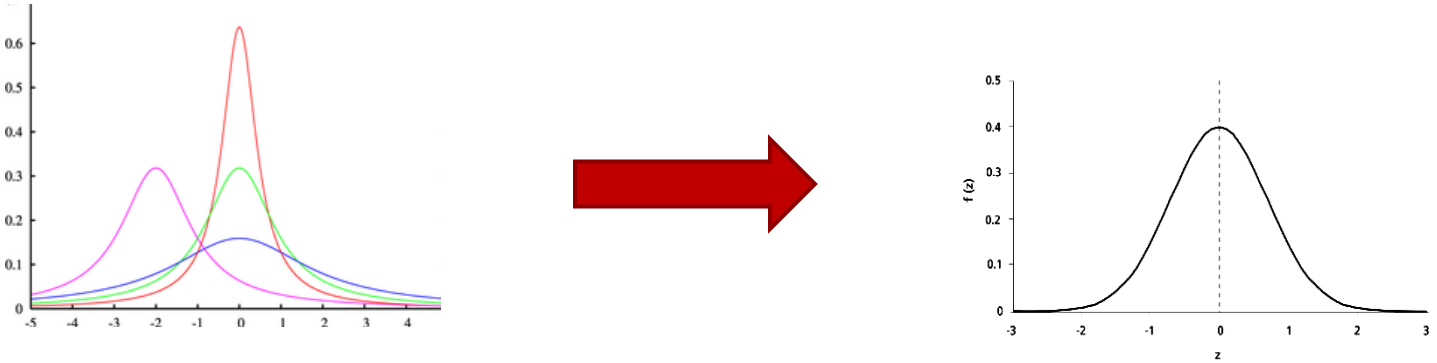
- The case of multiple experts is important
- When elicitation is used to provide expert input to a decision problem with substantial consequences, we generally want to use the skill of as many experts as possible
- However, they are not separate pieces of evidence
  - The experts typically base their judgements on the same or similar information
  - It is better to treat them as different attempts to formulate the same information
    - The prior information
- The question then arises of how to pool their judgements into a single prior density

# Aggregating expert judgements

- Two approaches
  - Aggregate the distributions
    - Elicit a distribution from each expert separately
    - Combine them using a suitable formula
    - Called 'mathematical aggregation' or 'pooling'
  - Aggregate the experts
    - Get the experts together and elicit a single distribution
    - Called 'behavioural aggregation'
- Neither is without problems

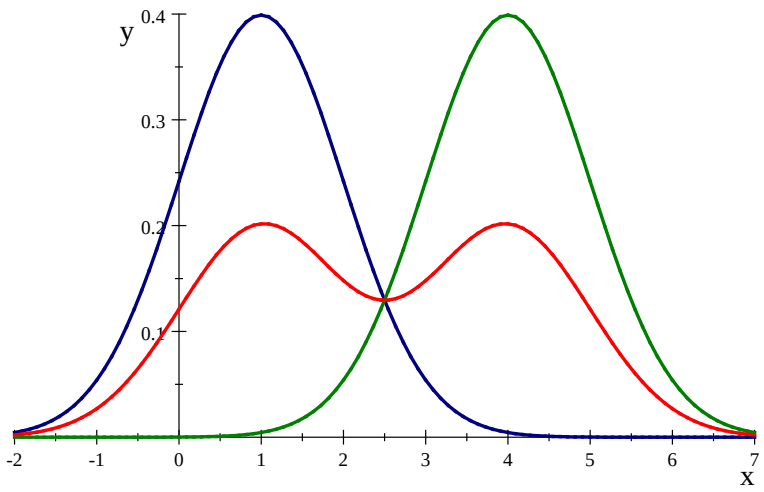
# Opinion pooling

- Mathematical aggregation methods combine several distributions into one
- Notation:
  - N experts
  - Their distributions are $f_i(x)$, for i = 1, 2, ..., N
  - Combined distribution $f_0(x)$



# Linear opinion pool

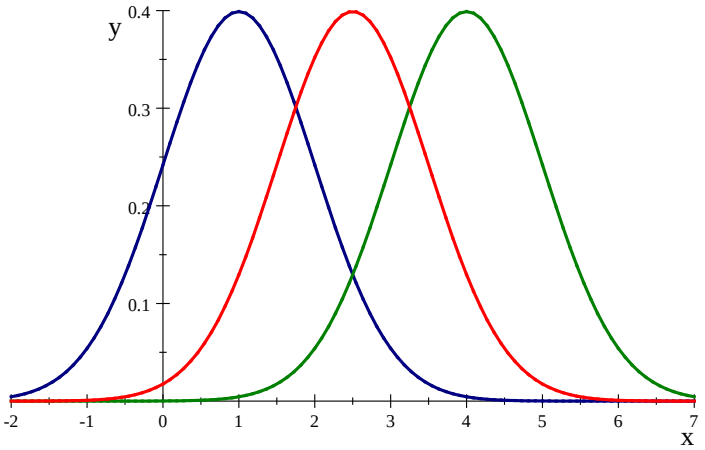
- The simplest way to combine the experts' probability distributions is to average them
  - $f_0(x) = \Sigma_i f_i(x) / N$
  - Blue and green lines = two expert distributions
  - Red line = pooled distribution




- More generally, we can give weights $w_i$ to the experts
  - $f_0(x) = \Sigma_i w_i f_i(x) / \Sigma_i w_i$
  - To reflect the experts' different levels of expertise

# Multiplicative opinion pool

- If instead we average the logarithms of the experts' probability distributions this amounts to multiplying
  - $f_0(x) = (\Pi_i f_i(x))^{1/N}$
    - Then scale the result so that it integrates to 1
  - Blue and green lines = two expert distributions
  - Red line = pooled distribution

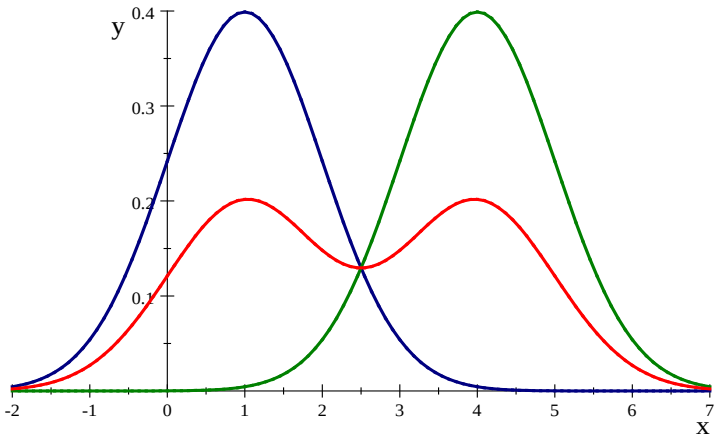



- More generally, we can give weights $w_i$ to the experts
  - $f_0(x) = (\Pi_i f_i(x)\, w_i)^{1/\Sigma_i w_i}$
    - Then scale the result so that it integrates to 1
  - To reflect the experts' different levels of expertise

# Practical comparison

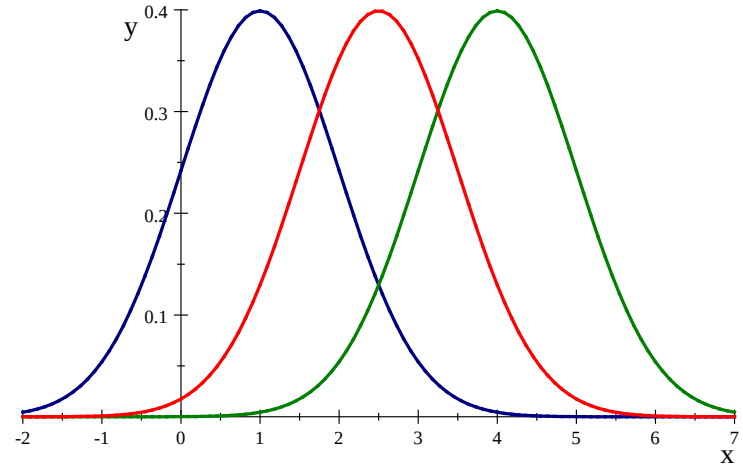
## Linear

- Each expert could be right
  - Pooled distribution covers the range of their beliefs




## Multiplicative

- Both experts are right
  - Pooled distribution based on the intersection of beliefs



# Theoretical comparison

**Linear**

- Consistent when you simplify
  - $Pr_0(X > 0)$ is average of $Pr_i(X > 0)$
- But not when you add information
  - $f_0(x \mid X > 0)$ is not the average of $f_i(x \mid X > 0)$

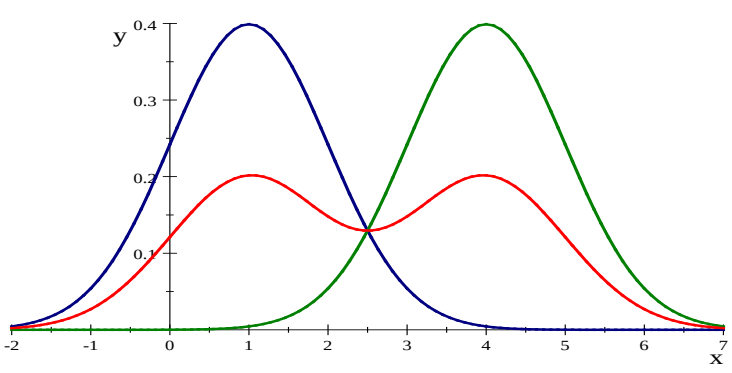
**Multiplicative**

- Consistent when you add information
  - $f_0(x \mid X > 0)$ is the log-average of $f_i(x \mid X > 0)$
- But not when you simplify
  - $Pr_0(X > 0)$ is not the log-average of $Pr_i(X > 0)$

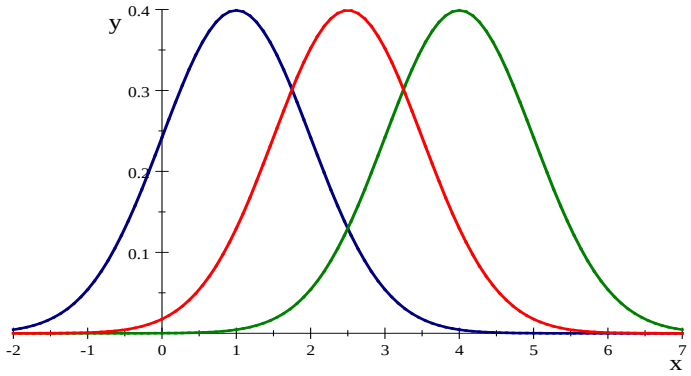
There is no way to pool without losing some desirable properties

# Behavioural aggregation

- The alternative to mathematical aggregation
- Get the experts in a room together and ask them to elicit a single distribution
- Advantages
  - Opportunity to share knowledge
  - Avoids arbitrary choice of pooling rule
  - Allows more subtle forms of aggregation




?



# But …

- More psychological hazards
  - Group dynamic – dominant/reticent experts
  - Tendency to end up more confident
  - Block votes
- Requires careful management
- What to do if they can't agree?
  - End up with two or more composite distributions
  - Need to apply mathematical pooling to these
  - But this is rare in practice

# Meta-analysis

Synthesising indirect evidence

# Basic normal meta-analysis

- The canonical context is to synthesise evidence from a set of clinical trials, all testing the same drug/treatment
  - These are usually found from a systematic review of the literature
- In general, suppose we have found N trials
  - Denote evidence from trial t by $E_t$
    - Could in principle be individual patient data
    - More usually it comprises summary statistics
  - Consider the case where $E_t = (z_t, s_t)$
    - Where $z_t$ is estimated treatment effect and $s_t$ its standard error
  - And we assume normally distributed estimate
    - Lots of other situations arise
  - We want to synthesise to make inference about true effect $\theta$

# A simple model

- Assume $z_t \sim N(\theta, s_t^2)$
  - Meaning $z_t$ has a normal distribution with mean the true effect $\theta$ and with variance $s_t^2$
- These are now multiple data items as in Session 1
  - Combine with prior information
    - Usually formulated as weak prior
  - Derive posterior inferences about $\theta$
- However, this may be too simple
  - Experience shows trial estimates are often too varied
  - Each trial has different features
    - Recruitment criteria, trial conditions
    - So each may be estimating a slightly different $\theta$

# Model with trial random effects

- A more sophisticated model is $z_t \sim N(\theta + \varepsilon_t, s_t^2)$
  - Where $\varepsilon_t$ is an effect due to the specific features of trial t
  - And assume $\varepsilon_t \sim N(0, \tau^2)$
    - Where $\tau^2$ is the variance of the trial effects, an unknown parameter
  - Complete the Bayesian formulation with (weak) priors for $\theta, \tau^2$
    - But genuine prior information about $\tau^2$ would be useful if available
- Equivalent expressions of this model
  - $z_t \sim N(\theta, s_t^2 + \tau^2)$
    - Multiple data items but with enhanced variance
  - $z_t \sim N(\mu_t, s_t^2)$, $\mu_t \sim N(\theta, \tau^2)$
    - A more complex Bayesian analysis with extra parameters $\mu_t$
    - The original expression is similar with extra parameters $\varepsilon_t$
    - Both allow estimation of the individual trial effects

# Numerical example

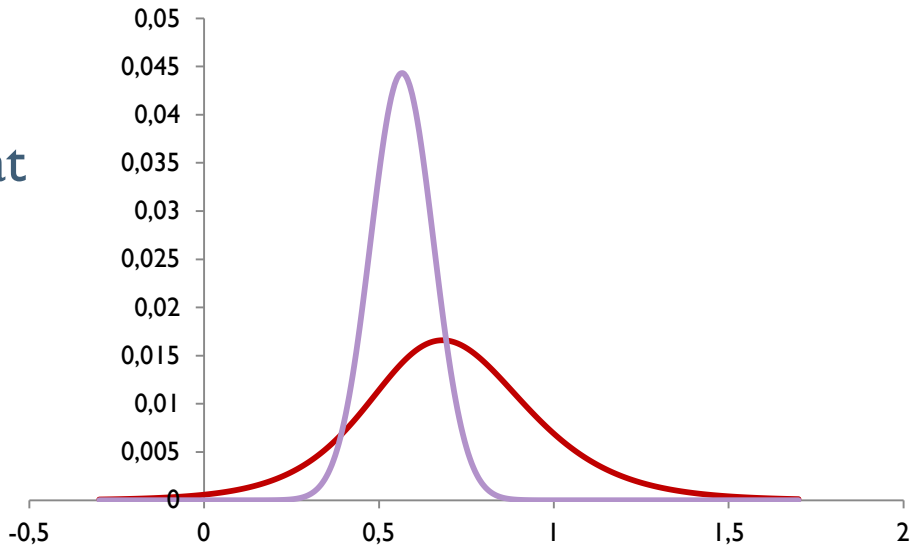
- The data here are made up, but intended to represent a realistic scenario
  - 9 trials of widely varying sizes (and hence standard errors)

| Trial | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| $z_t$ | 1.72 | 0.66 | 0.95 | 1.01 | -0.56 | 0.07 | 1.60 | 0.38 | 0.97 |
| $s_t$ | 0.55 | 0.54 | 0.35 | 0.25 | 0.53 | 0.27 | 0.74 | 0.13 | 0.27 |

  - If all the trials are providing unbiased estimates of the same $\theta$ then for instance the difference between trial 6 and trials 4 and 9 is surprising

# Numerical example (contd.)

- The simple model produces a normal posterior distribution for $\theta$ with mean 0.566 and standard deviation 0.090
- The random effects model produces a posterior distribution for $\theta$ with mean 0.703 and standard deviation 0.270
  - The posterior mean of $\tau^2$ is 0.52
- The two results are quite different
  - The value of $\tau^2$ indicates that treatment effect varies widely with patient group
- Even when random effects appear small they should not be ignored



# Data gaps

- It is very common to find that data concern a slightly different situation from the one of interest
  - We are interested in a parameter $\theta$, but the data inform us about a different parameter $\alpha$
  - $\alpha$ is related to $\theta$, but we need to link the two
    - in order to use the data to learn about $\theta$
- I call this a data gap
  - Or data discrepancy
- In the case of random effects meta-analysis
  - Each trial provides information about a different parameter $\alpha_t$
  - We build the link with $\theta$ by writing $\alpha_t = \theta + \varepsilon_t$
    - Together with a distribution for the discrepancy $\varepsilon_t$

# Data gaps (contd.)

- There is usually very little information about discrepancies
  - In this case just 9 trials to learn about the discrepancy variance
  - Sometimes there is essentially no data to fill data gaps
    - Hence the name
    - Expert elicitation is then the only feasible way to fill them
- We'll encounter similar gaps and discrepancies elsewhere in this course

# More complex meta-analyses

Mixed treatment comparisons and meta-regression

# Beyond simple treatment comparisons

- Basic meta-analysis combines data from several trials about a single treatment effect
  - Usually a difference between the treatment of interest and a comparator
    - An alternative treatment, placebo or standard care
- Often we don't find multiple trials all with the same comparator
  - Some compared with placebo and some compared with active comparator
  - Combination therapies
  - Here are some simple examples of what is called Mixed Treatment Comparison

# Different comparators

- We are interested in the effect of treatment A compared with treatment B
  - We have some trials comparing A with C and some comparing B with C
    - Model the effects in the second group as $\beta + \varepsilon_t$
    - Model the effects in the first group as $\theta + \beta + \varepsilon_t$
    - Then $\theta$ represents the additional effect of A versus B
  - Now suppose we also have a 3-arm trial comparing A, B and C
    - Now need to model each arm of each trial separately
    - For a treatment A arm we write the mean response as $\theta + \beta + \eta_a + \varepsilon_t$
      - Where $\eta_a$ is a random effect of arm a and $\varepsilon_t$ is a trial random effect as before

# Combination therapies

- In cancer treatment, it is normal to give patients a combination of drugs
  - Also found in other serious conditions
    - E.g. patients who have had a stroke or heart attack
  - We may want to estimate the effect of a combination which has never appeared yet in any trial
  - Suppose in a treatment arm, patients receive drugs A, B and C
    - Then the mean effect in that arm could be modelled as a sum of drug effects $\theta_A + \theta_B + \theta_C$ plus random effects
    - But this assumes independent effects, and some interaction effects may also be modelled
    - Interactions in combinations that have not yet been tested represent a data gap for which there is no evidence

# Introducing covariates

- These mixed treatment comparison models are particular cases of a more general meta-regression framework
  - For each trial effect or treatment arm we may have some covariates that might explain trial differences
    - Average patient age or disease severity
    - Trial duration
  - We can model the mean response on a treatment arm or the mean effect in a trial using conventional regression techniques
    - E.g. $\theta + \beta s_t + \varepsilon_t$
    - Where $s_t$ is average severity in trial t and $\beta$ is regression coefficient
  - Mixed treatment comparisons involve 0-1 covariates
    - 0 if treatment not used, 1 if it is used

# Summary of session 2

- Combining beliefs of several experts is a different kind of synthesis
  - Mathematical aggregation is simple but involves an arbitrary choice of pooling rule
  - Behavioural aggregation is potentially more powerful but also more complex
    - SHELF system to help facilitators
- Meta-analysis is a powerful range of techniques for synthesising published data sources
  - Widely used for synthesising evidence from clinical trials
    - Including mixed treatment comparisons and meta-regression
  - Can be adapted to many other contexts
- We frequently need to be aware of data gaps/discrepancies
  - Gaps are often hard to fill and may require expert elicitation