Day 4 Bayesian statistics

Recap of Day 3. Frequentist statistics

- Focus on type-I error rate: $\alpha = P(\text{reject null} \mid \text{null true})$
- Significance: unlikely event, from point of view of null statistics.
- Involves calculating probability of statistics given a hypothesis.
- Machinery can be applied to many different tests.
- Widespread usage, particularly in biological/medical literature.

4.1 Bayesian approach

4.1.1 Bayesian basics

- Reminder: Bayes theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Apply this to data D and a hypothesis H.
  Allows to update beliefs using information from experiment:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} \quad \text{or} \quad P(H|D) \propto P(D|H)P(H)$$

- $P(H)$ is prior belief on the hypothesis H
- $P(D|H)$ is called the likelihood: probability data given hypothesis.
- $P(D)$ is probability of data - a normalisation constant.

4.1.2 Likelihood

- For independent data points, update as a product.

$$P(H|\{D_k\}) \propto \left[\prod_k P(D_k|H)\right] P(H)$$

- Not always simple to calculate the likelihood $P(D_k|H)$

- If there are two hypothesis H_1, H_2 then we can normalise this as follows

$$P(H_1|\{D_k\}) = \frac{1}{\mathcal{N}}\left[\prod_k P(D_k|H_1)\right] P(H_1)$$
$$P(H_2|\{D_k\}) = \frac{1}{\mathcal{N}}\left[\prod_k P(D_k|H_2)\right] P(H_2)$$

- where
  $\mathcal{N} = \left[\prod_k P(D_k|H_1)\right] P(H_1) + \left[\prod_k P(D_k|H_2)\right] P(H_2)$

4.1.3 Discrete parameter example

- Consider a game where one of two coins is used:
  one is fair $p_1 = 0.5$ the other is biased for heads $p_2 = 0.7$
- From experience the biased coin is used 20$\%$ of the time.
- We will associate the hypothesis C_1 and C_2 with the two coins.
- A coin is chosen and you see the following sequence: h, t, h, h, h, h, t, h
- **What is the prior distribution?**
  $P(C_1) = 0.8$ and $P(C_2) = 0.2$
- **What are the two likelihoods for the first flip?**
  $P(\text{h}|C) = p$ so that $P(\text{h}|C_1) = 0.5$ and $P(\text{h}|C_2) = 0.7.$
- **What is the posterior distribution after the first flip?**
  $P(C_1|\text{h})= \frac{0.5\times0.8}{0.5\times0.8+0.7\times0.2} = 0.741$ and
  $P(C_1|\text{h})= \frac{0.7\times0.2}{0.5\times0.8+0.7\times0.2} = 0.259$
- **Question.** What is the posterior distribution at the end of the sequence?
- **Question.** What would be the distribution if you had no prior knowledge?

4.1.4 Continuous parameter example

- A coin is flipped N times with n heads.

- What can we infer about it's bias? Mean gives $\hat{p} = n/N$. What else?

- Let's consider that the unknown parameter p is like a random variable.

- Let $f_0(p)$ be the distribution of our prior view on p.
- Let $\mathcal{L}(D|p)$ be the likelihood (probability of data given p).
- Let $f(p|D)$ be the posterior distribution on p.

- From Bayes $f(p|D) = \frac{1}{\mathcal{N}}\mathcal{L}(D|p)f_0(p)$ where $\mathcal{N} = \int_0^1 dp'\mathcal{L}(D|p')f_0(p')$.

- **Question.** What is $\mathcal{L}(D|p)$?

- **Question.** For a coin with a 0.7 heads bias generate 20 coin flips and
  update your prior. Use two priors: one uniform the other $\propto p(1-p)$.
  [HINT] Discretise the range of p and approximate the integrals as sums.

4.2 Bayesian versus frequentist

4.2.1 Berger and Berry (1988)

- Frequentist approach often justified as avoiding subjective/unknown prior

- Influencial paper (500+ citations) challenging objectivity of frequentist approach http://www2.hawaii.edu/~cbaajwe/Ph.D.Seminar/BergerandBerry1988.pdf (http://www2.hawaii.edu/~cbaajwe/Ph.D.Seminar/BergerandBerry1988.pdf)

The experiment

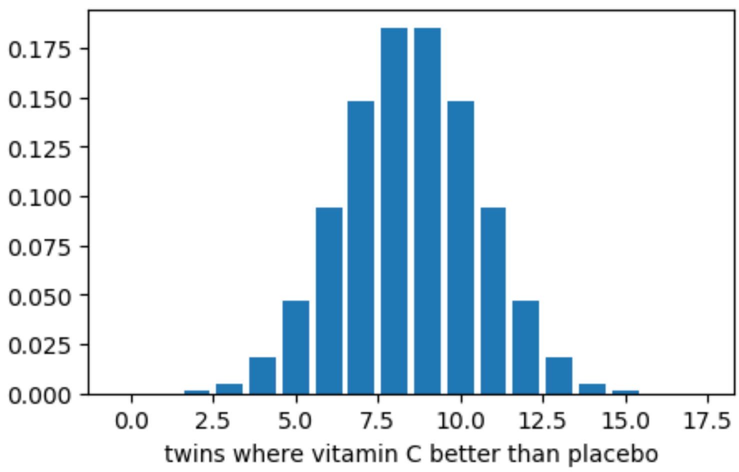
- Test if vitamin C (C) relieves cold symptoms against a placebo (P)
- 17 identical twins, one gets C the other P
- Results: 13 taking C get better, 4 taking P get better.
- What is the significance of this result?

- Need to calculate the p-value and compare with, say 5% or 1% levels.
- What is the distribution under the null hypothesis (no effect of vitamin C)?

$$P(C = k|H_0) = \binom{17}{k}(0.5)^k$$

In [6]:

```
k=collect(0:17)
Pk=pdf.(Binomial(17,0.5),k);
figure(figsize=(5,3));
bar(k,Pk); xlabel("twins where vitamin C better than placebo");
```



- The result was 13, so as extreme is $k = 0, 1, 2, 3, 4$ and $13, 14, 15, 16, 17.$
- The p-value is $2\sum_{k=0}^{4} P_k = 0.049$ and just within the 5% significance range.

4.2.2 p-value depends on the intent

- This seems objective, but what if another experiment was envisaged?

- **For example:**
  Stop when you have at least 4 where C is better and 4 where P is better.

- Imagine the 4th P occured on the 17th trial, generating an identical data set.

- The null distribution will be very different.

- Hopefully it won't affect the p-value...

- **Question.** Check the p-value for this experiment by generating artifical data.
  The statistic now is the number of trials (twins).

4.3 Additional topics

4.3.1 Conjugate priors
4.3.2 Credible intervals

4.3.1 Conjugate priors

- For the coin with the unknown bias we discretized the space of p.
- This might get numerically expensive with many variables (bins).
- Numerical integration can be avoided if there is a congugate prior.
- Allows for a low-dimensional parameterisation of a continuous prior.

Example

- The conjugate prior to the Binomial distribution is a Beta distribution $p_{\alpha\beta}$
  where $p_{\alpha\beta} = p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta)$ and
  $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$

- Imagine we start with a Beta-distribution prior parameterised by α and β.
- **Question:** Show that if a coin is flipped $n_h + n_t$ times with n_h heads and n_t tails,
  the posterior is a Beta distribution with $\alpha' = \alpha + n_t$ and $\beta' = \beta + n_h$.

4.3.2 Credible intervals

- These are analagous to the frequentist confidence intervals.

- They have a more direct interpretation using the posterior distribution.

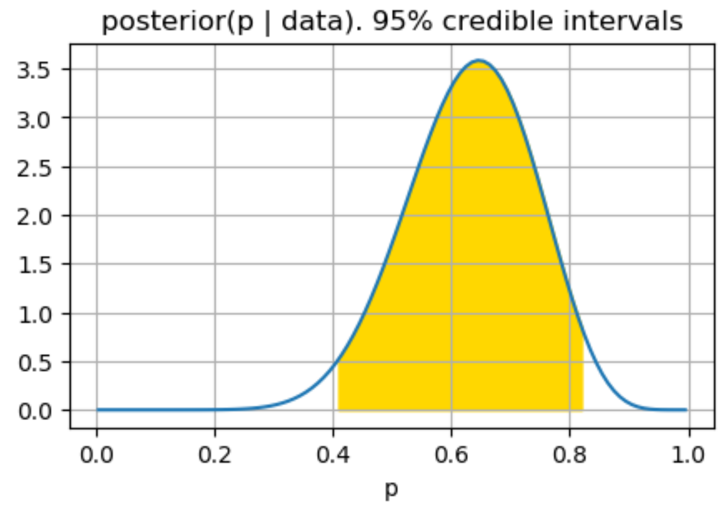
- There is some choice. A 95% credible interval could be:
  (1) The shortest interval containing 95% of the density
  (2) A symmetric distribution, with 2.5% discounted on either side.

- Example for the coin with the unknown bias (using a Beta distribution).

In [8]:
```
dp=0.01
p=collect(dp/2:dp:1.0-dp/2)
a,b=12,7

f=pdf.(Beta(a,b),p);
p1=invlogcdf.(Beta(a,b),log(0.025))
p2=invlogcdf.(Beta(a,b),log(1-0.025))
pp=p1:dp:p2
ff=pdf.(Beta(a,b),pp)

figure(figsize=(5,3))
plot(p,f); xlabel("p"); title("posterior(p | data). 95% credible intervals")
fill_between(pp,0,ff,color="gold"); grid();
```



4.4 Summary and additional questions

Day 4 Bayesian statistics

Questions

Make sure you have understood and done all the questions in the lectures.

The questions below are to be handed in for marking on the 19th October 2017

Q4.1 Maximum likelihood: normal distribution
Q4.2 Posterior for normal distribution

Q4.1 Maximum likelihood: normal distribution

- Consider a sample $X_1 \cdots X_n$ from a normal distribution.
- What is the likelihood they were drawn from a distribution with parameters
  mean μ and variance σ^2.
- What values $\hat{\mu}$ and $\hat{\sigma}^2$ maximise this likelihood?
- Is $\hat{\sigma}^2$ a biased or unbiased estimator?

Q4.2 Posterior for normal distribution

- For some choice of μ and σ^2, generate $n = 5$ normally distributed random numbers.
- Using a flat prior, calculate the posterior density $p(\mu, \sigma | X_1 \cdots X_n)$.
- Note this is a two-dimensional density. You will have to create a 2D grid of data points.
- Plot the 2D distribution.
- Calculate the maximum-likelihood parameters $\hat{\mu}$ and $\hat{\sigma}^2$ using your formula from Q4.1. Confirm that their coordinates coincide with the peak of the posterior distribution.
- Use the posterior distribution to calculate the marginal distributions for μ and σ and plot these. Do their peaks also coincide with $\hat{\mu}$ and $\hat{\sigma}^2$?