

1. MLE vs MAP

- (a) You have a bag that contains red and blue marbles, but you don't know how many of each color are in the bag. Suppose you sample 5 marbles with replacement from the bag and obtain the results R, B, R, R, R . What is the likelihood of this draw? Give your answer as a function of θ , the fraction of blue marbles.

Solution: The likelihood is

$$p(\text{sample}|\theta) = \prod_{i=1}^5 p(\text{Observation}_i|\theta) = (1 - \theta) \cdot \theta \cdot (1 - \theta)^3 = \theta \cdot (1 - \theta)^4$$

- (b) Let $L(\theta)$ be the function from part (a). What is the name for the quantity $\arg \max_{\theta} L(\theta)$?

Solution: MLE

- (c) Now suppose you have a prior over θ . In particular, you believe that θ was drawn from a $\text{Beta}(\alpha, \beta)$ distribution. What is the posterior over θ given the observed samples in part (a)?

(**Hint:** Use conjugate prior rules.) **Solution:** We have a beta prior and a Bernoulli likelihood. Since beta and Bernoulli are conjugate, we can use the update rule from lecture. The posterior is $\text{Beta}(\alpha + 1, \beta + 4)$. Specifically,

$$\begin{aligned} p(\theta|x) &\propto p(x|\theta)p(\theta) \\ &\propto \left(\theta^{\text{num blues}} (1 - \theta)^{\text{num reds}} \right) \theta^{\alpha-1} (1 - \theta)^{\beta-1}. \end{aligned}$$

So

$$\begin{aligned} \theta|x &\sim \text{Beta}(\alpha + \text{num blues}, \beta + \text{num reds}) \\ &= \text{Beta}(\alpha + 1, \beta + 4). \end{aligned}$$

- (d) Place the following quantities in order from smallest to largest:

- The MLE for θ
- The MAP under a $\text{Beta}(1, 100)$ prior.
- The MAP under a $\text{Beta}(100, 1)$ prior.

Solution: The $\text{Beta}(1, 100)$ prior will skew the posterior to be close to 0, while $\text{Beta}(100, 1)$ will skew it to be close to 1. Meanwhile, the MLE is somewhere in-between (in fact, equal to $\frac{1}{5}$ in this case – just differentiate likelihood or log-likelihood w.r.t θ and set to 0). So smallest to largest: b,a,c.

2. Say you've observed a sequence of coin flips, X_1, \dots, X_n , all using the same coin, which has some probability of landing heads, θ_h . The total number of heads you observe, H , follows a binomial distribution with probability mass function

$$p(H = k|\theta_h) = \binom{n}{k} \theta_h^k (1 - \theta_h)^{n-k}.$$

Suppose we want to do Bayesian inference, with a prior $p(\theta_h)$. Since $\theta_h \in [0, 1]$, we'll take our prior distribution to be $\text{Beta}(\alpha, \beta)$.

What is the posterior distribution for θ_h , in terms of α , β , n , and k ? What is the corresponding LMSE estimate for θ_h ? (*Hint*: it turns out that the beta distribution is a conjugate prior for the binomial distribution.)

Solution:

$$\begin{aligned} p(k|\theta_h) \cdot p(\theta_h; \alpha, \beta) &= \binom{n}{k} \theta_h^k (1 - \theta_h)^{n-k} \cdot \frac{(\alpha + \beta - 1)!}{(\alpha - 1)! (\beta - 1)!} \theta_h^{\alpha-1} (1 - \theta_h)^{\beta-1} \\ &= \frac{n!}{(n-k)! k!} \cdot \frac{(\alpha + \beta - 1)!}{(\alpha - 1)! (\beta - 1)!} (\theta_h)^{k+\alpha-1} (1 - \theta_h)^{n-k+\beta-1} \end{aligned}$$

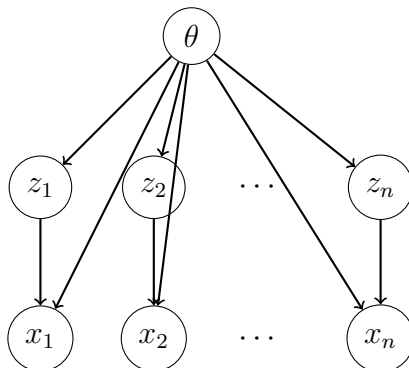
Where the second line follows from rearranging terms. In terms of θ_h , the first two fractions are constant, so we can write

$$\begin{aligned} p(\theta_h|H = k) &\propto f(k|\theta_h) \cdot p(\theta_h; \alpha, \beta) \\ &\propto (\theta_h)^{k+\alpha-1} (1 - \theta_h)^{n-k+\beta-1} \end{aligned}$$

From this we conclude that the posterior distribution has a beta distribution with shape parameters $(k + \alpha)$ and $(n - k + \beta)$.

LMSE estimate of θ : $\theta_{\text{LMSE}} = \mathbb{E}[\theta_h|H = k] = \frac{k+\alpha}{n+\beta+\alpha}$.

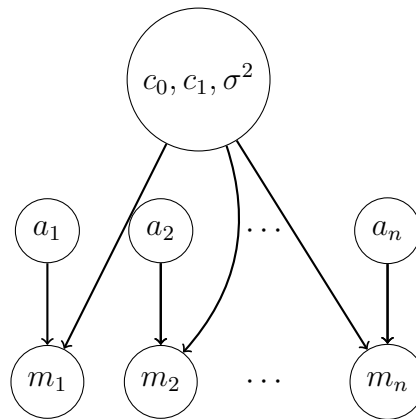
3. Last lecture, we were introduced to *graphical models*, which are flexible diagrams to express the relationships between random variables. An important special case is a Bayesian hierarchical model (see figure below for an example). In a Bayesian hierarchical model, observations X are independent given the latent variables Z , and each observed variable depends only on its corresponding latent variable and a set of *hyperparameters* θ :



As an example, suppose that we are running a marine observatory and have information about the mass m_i of each seal i , for $i = 1, \dots, n$. We believe that the mass of a seal should be approximately normally distributed with variance σ^2 and mean equal to $c_1 \cdot a_i + c_0$, where a_i is the unknown age of seal i and c_0, c_1, σ^2 are unknown parameters (but do not depend on i). We know that the ages are drawn uniformly from 0 to 35.

Draw a graphical model for this situation, similar to the Bayesian hierarchical model above. What are the observed variables? What are the latent variables? What are the hyperparameters?

Solution: Observed: m_i ; Latent: a_i ; Hyperparameters: c_0, c_1, σ^2 **Solution:**



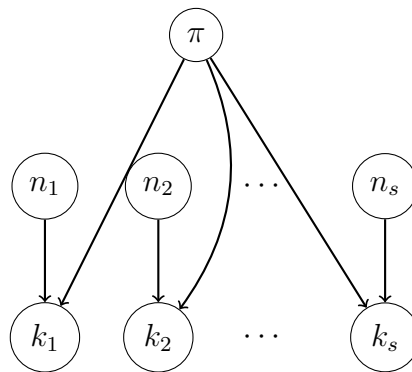
4. **COVID Meta-analysis.** We will next explore a graphical model that was used in a research paper on COVID-19¹. The specific question was: *How likely is someone with COVID-19 to transmit it to someone else in the same household?*

For this question, there had been a number of studies, but most studies were with only a small number of participants and there was significant variability in terms of study conditions (baseline prevalence of COVID-19, degree of local restrictions, etc.). We therefore want to do a meta-analysis to develop a single more robust estimate of the probability of household transmission.

Suppose that there are s studies, and that study i has n_i “participants” (household members of COVID-19 patients), and that k_i of these participants eventually tested positive for COVID-19. We are interested in π , the probability that a participant gets COVID-19.

We will consider several increasingly complex ways of modeling this problem. For each of these, we will provide the nodes of the graphical model, and you should shade in the nodes that are observed and draw the relevant arrows.

- (a) In the simplest model, we assume that every person in every study is infected independently with the same probability π (π is called the “secondary attack rate”).



What is the conditional probability distribution for k_i given its parent nodes?

Solution:

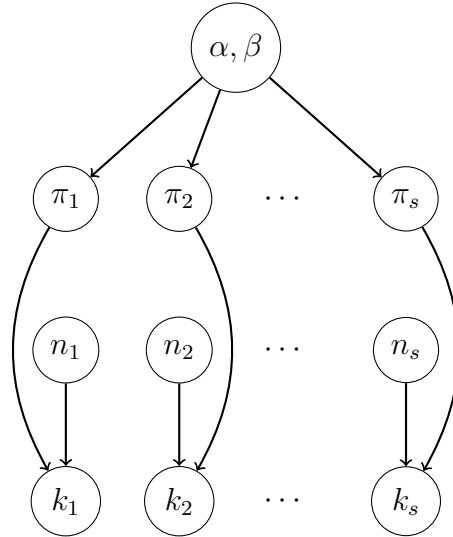
$$k_i | n_i, \pi \sim \text{Binomial}(n_i, \pi)$$

- (b) There are unfortunately some pretty big issues with the model above. With your partner, brainstorm possible issues with the model as given.

Solution: Different study populations (different vaccination rates, population densities), so want different π_i 's.

¹<https://www.medrxiv.org/content/10.1101/2020.05.23.20111559v2>

- (c) We'll next address one issue with the model above—that each study i actually has its own unique secondary attack rate π_i , due to variability in study conditions. To tie together the different π_i , we'll place a *prior* over the π_i , assuming that $\pi_i \sim \text{Beta}(\alpha, \beta)$ independently for each i .

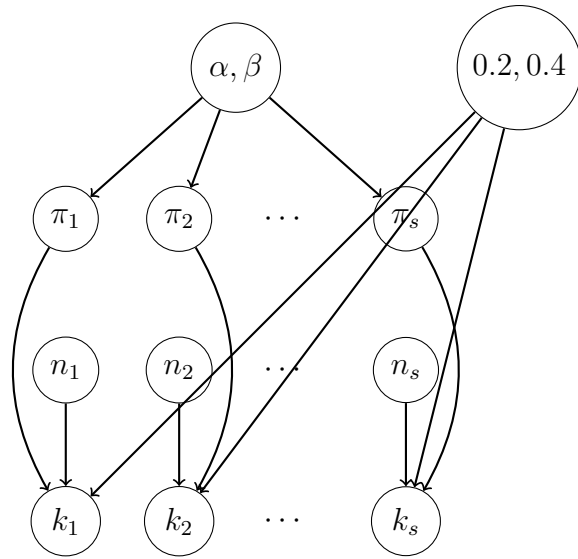


- (d) What does the quantity $\frac{\alpha}{\alpha + \beta}$ represent in this model? What about $\alpha + \beta$?

Solution: $\frac{\alpha}{\alpha + \beta}$: expected value of secondary attack rate.

$\alpha + \beta$: rough measure of inverse variance of π_i . Think about the shape of a beta distribution. For α, β such that $\alpha + \beta$ is large, π_i will have smaller variance. For α, β such that $\alpha + \beta$ is small, π_i will have larger variance.

- (e) Finally, suppose that we want to account for the fact that COVID tests can have false negatives, so that someone with COVID-19 only tests positive with probability $1 - f$, for some value f that denotes the false negative rate. Suppose that we do not know the FNR but have a prior for f that is uniformly distributed between 0.2 and 0.4. Modify the graphical model above by adding appropriate nodes and arrows to account for this. What is the new conditional distribution for k_i given its parent nodes?



Solution:

$$k_i | n_i \sim \text{Unif}(0.2n_i, 0.4n_i)$$

$$k_i | n_i, \pi_i \sim \text{Binomial}(n_i, \pi_i)$$

5. (Challenge Question)

- (a) Consider a likelihood function $p(X | \mu) = \frac{1}{\sqrt{2\pi}} \exp(-(X - \mu)^2/2)$, which is normally-distributed with mean μ and variance 1. Show the conjugate prior is also a normal distribution. What is the corresponding update rule?
- (b) Now suppose the likelihood function is $p(X | \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-X^2/2\sigma^2)$, a normal distribution with mean 0 and variance σ^2 . Find a distribution family that serves as a conjugate prior, and determine the corresponding update rule.
- (c) Finally, consider the general case $p(X | \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(X - \mu)^2/2\sigma^2)$. Find a valid conjugate prior and update rule in this case.

Feedback Form

On a scale of 1-5, where 1 = much too slow and 5 = much too fast, how was the pace of the discussion section?

1 2 3 4 5

Which problem(s) did you find most useful?

Which were least useful?

Any other feedback?