(1) A Bayesian model

- A Bayesian statistical model is a joint distribution of the data x and the parameters θ
  - So, the model parameters have a distribution, too!
- Define the prior distribution as $p(\theta)$
- Define the likelihood as $p(x|\theta)$
- Then, the Bayesian model is the following joint distribution:

$$p(x,\theta) = \overbrace{p(x|\theta)}^{likelihood}\ \overbrace{p(\theta)}^{prior}$$

Bayesian inference

- Problem: What is the distribution of the parameters, after conditioning on the observed data?
- Solution: derive the posterior distribution of the parameters by applying the Bayes' formula:

$$\overbrace{p(\theta|x)}^{posterior} = \frac{p(x,\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta)$$

Example: Binomial likelihood

- Coin tossing with probability of success π; observe X successes out of N trials
  - Prior: $\pi \sim \text{Beta}(\alpha, \beta)$
  - Likelihood: $X \sim \text{Binomial}(N, \pi)$
- The posterior of π:

$$p(\pi|X) \propto \pi^X (1-\pi)^{N-X} \times \frac{\pi^{\alpha-1}(1-\pi)^{\beta-1}}{B(\alpha,\beta)}$$

- This means that $\pi|x \sim \text{Beta}(\alpha + X, \beta + N - X)$

Conjugate prior distribution

- A class of prior distributions $p(\theta)$ is *conjugate* for a given family of likelihood functions, if the resulting posterior distributions belong to same class as $p(\theta)$
- Example: Beta prior is conjugate to Binomial likelihood (see above)
- Example: Gamma prior is conjugate to Poisson likelihood (next slide)

Example: Poisson likelihood

- Assume a Poisson likelihood:

$$p(N|\theta) = \frac{(\theta Y)^N \exp(-\theta Y)}{N!}$$

- Assume a Gamma prior: $\theta \sim \text{Gamma}(\alpha, \beta)$
- The posterior of parameter θ:

$$p(\theta|N) \propto \frac{\theta^N \exp(-\theta Y)}{N!} \times \theta^{\alpha-1} \exp(-\beta\theta)$$

- This means that $\theta|N \sim \text{Gamma}(\alpha + N, \beta + Y)$

Survival likelihood

- Survival data $(T_i, D_i),\ i = 1, \ldots, N$
- Survival likelihood:

$$p(\{T_i, D_i\}|\theta; \{Z_i\}) =$$

$$\left(\prod_{i=1}^N \lambda(t; \theta, Z_i)^{D_i}\right) \times \exp\left(-\sum_{i=1}^N \int_0^\infty Y_i(u)\lambda(u; \theta, Z_i)\right)$$

- This is also called a Poisson likelihood
- Gamma priors for hazard rates are convenient
  - Use of conjugate priors sometimes feasible

Computations in practice

- In practice, posterior distributions are often explored through iterative sampling from (full) conditional posterior distributions
- For example, for two parameters θ and ϕ, consider conditional distributions

  $p(\theta|\phi, x)$ and $p(\phi|\theta, x)$

- Sometimes it is possible to derive these conditionals explicitly
  - Gibbs sampling (see example later)
- If not possible, use other numerical sampling algorithms (e.g. Metropolis-Hastings)

(2) Non-parametric smoothing

- Assume a piece-wise constant model for the incidence rate in K time bands: $\lambda_j, \ j = 1, \ldots, K$
- The data $(N_j, Y_j), \ j = 1, \ldots, K$ (number of cases + person-time)
- Assume a Markovian prior for λ_j:

  $$\log \lambda_j | \log \lambda_{j-1} \sim \text{Normal}(\log(\lambda_{j-1}), \sigma^2)$$

- When $\sigma \to 0$, all λ_j's $\to$ constant (i.e. maximal smoothing). When $\sigma \to \infty$, independent $\lambda_j' s$ (i.e. no smoothing at all)