Lecture Summary

Another Example of Bayesian estimation - Normal distribution

- Let $X_1, \ldots, X_n$ be a random sample from $N\left(\theta, \sigma^2\right)$ where σ^2 is known
- Let the prior distribution of θ be $N\left(\mu_0, \nu_0^2\right)$ where μ_0 and ν_0^2 are known.
- Show that the posterior distribution $p(\theta \mid \mathrm{x})$ is $N\left(\mu_1, \nu_1^2\right)$ where

$$\mu_1 = \frac{\sigma^2\mu_0 + n\nu_0^2\bar{\mathrm{x}}_n}{\sigma^2 + n\nu_0^2} \quad \text{and} \quad \nu_1^2 = \frac{\sigma^2\nu_0^2}{\sigma^2 + n\nu_0^2}$$

The posterior mean is a linear combination of the prior mean μ_0 and the observed sample mean.

Normal Distribution

- What happens when $\nu_0^2 \to \infty$?
- What happens when $\nu_0^2 \to 0$?
- What happens when $n \to \infty$?

Frame Title

Conjugate Priors

Let $X_1, X_2, \ldots$ be a random sample from $f(x \mid \theta)$. A family Ψ of distributions is called a conjugate family of prior distributions if for any prior distribution $p(\theta)$ in ψ the posterior distribution $p(\theta \mid \mathrm{x})$ is also in ψ

Conjugate priors

Likelihood	Prior	Posterior
$\text{Bern}(p)$	$\text{Beta}(\alpha, \beta)$	$\text{Beta}(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i)$
$\text{Pois}(\lambda)$	$\text{Gamma}(\alpha, \beta)$	$\text{Gamma}(\alpha + \sum_{i=1}^{n} x_i, \beta + n)$
$\text{Expo}(\lambda)$	$\text{Gamma}(\alpha, \beta)$	$\text{Gamma}(\alpha + n, \beta + \sum_{i=1}^{n} x_i)$
$\mathcal{N}(\theta, \sigma^2)$ known σ^2	$\mathcal{N}(\mu_0, \nu_0)$	$\mathcal{N}(\frac{\sigma^2\mu_0 + n\nu_0\overline{x_n}}{\sigma^2 + n\nu_0}, \frac{\sigma^2\nu_0^2}{\sigma^2 + n\nu_0^2})$

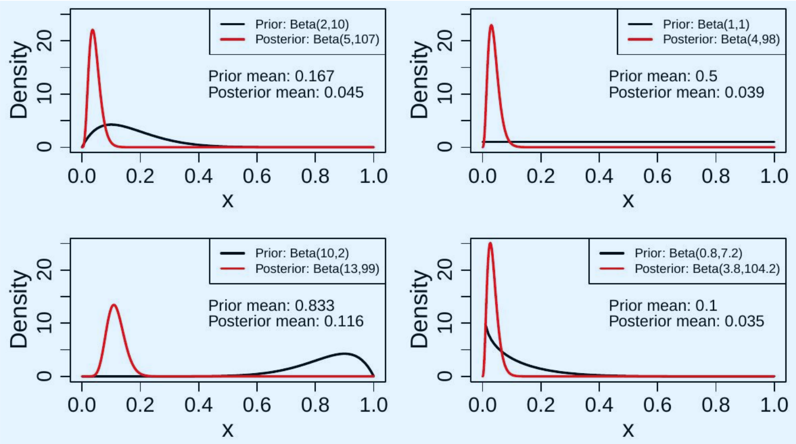
Prior distributions

- The prior distribution should reflect what we know apriori about θ
- For example: Beta$(2, 10)$ puts almost all of the density below 0.5 and has a mean $2/(2 + 10) = 0.167$, saying that a prevalence of more then 50% is very unlikely
- Using Beta $(1, 1)$, i.e. the Uniform $(0, 1)$ indicates that a priori all values between 0 and 1 are equally likely.

Choosing a prior

- Deciding what prior distribution to use can be very difficult
- We need a distribution (e.g. Beta) and its hyperparameters (e.g. α, β)
- When hyperparameters are difficult to interpret we can sometimes set a mean and a variance and solve for parameters
  E.g: What Beta prior has mean 0.1 and variance 0.1^2 ?
- If more than one option seems sensible, we perform sensitivity analysis

Sensitivity Analysis



We compare the posteriors we get when using the different priors.

Sensitivity Analysis

The posterior is influenced both by sample size and the prior variance

- Larger sample size $\Rightarrow$ less the prior influences the posterior
- Larger prior variance $\Rightarrow$ the less the prior influences the posterior Prior variance: 0.011

Improper priors

- Improper Prior: A "pdf" $p(\theta)$ where $\int p(\theta)d\theta = \infty$
- Used to try to put more emphasis on data and down play the prior
- Used when there is little or no prior information about θ.
- Not clear that an improper prior is necessarily "non-informative".
- Danger: We always need to check that the posterior pdf is proper! (Integrates to 1)

Point Estimator

- In principle, Bayesian inference is the posterior distribution.
- However, often people wish to estimate the unknown parameter θ with a single number.
- A statistic: Any function of observable random variables $X_1, \ldots, X_n$, $T = r(X_1, X_2, \ldots, X_n)$.
- Example: The sample mean $\bar{X}_n$ is a statistic

Definition (Estimator/Estimate)

Suppose our observable data $X_1, \ldots, X_n$ is i.i.d. $f(x \mid \theta), \theta \in \Omega \subset \mathbb{R}$.
- Estimator of θ : A real valued function $\delta(X_1, \ldots, X_n)$.
- Estimate of θ : $\delta(x_1, \ldots, x_n)$, i.e. estimator evaluated at the observed values.
- An estimator is a statistic and a random variable .

Loss Function

Loss function:

A real valued function $L(\theta, a)$ where $\theta \in \Omega$ and $a \in \mathbb{R}$.

$L(\theta, a) =$ what we loose by using a as an estimate when θ is the true value of the parameter.

Example Loss Functions

- Squared error loss function: $L(\theta, a) = (\theta - a)^2$
- Absolute error loss function: $L(\theta, a) = |\theta - a|$
- Zero-one loss: $L(\theta, a) = 0$, if $\theta = a$, 1, otherwise.

Bayes Estimator

Idea

Choose an estimator $\delta(X)$ so that we minimize the expected loss.

An estimator is called the Bayesian estimator of θ if for all possible observations x of X the expected loss is minimized. For given $X = x$ the expected loss is

$$E(L(\theta, a) \mid x) = \int_{\Omega} L(\theta, a) p(\theta \mid x) d\theta$$

Let $a^*(x)$ be the value of a where the minimum is obtained. Then $\delta^*(x) = a^*(x)$ is the Bayesian estimate of θ and $\delta^*(X)$ is the Bayesian estimator of θ.

Bayes Estimator

Theorem

The posterior mean $\delta^*(X) = E(\theta \mid X)$ *is the Bayes estimator for the squared error loss.*

$\min_a E(L(\theta, a) \mid x) = \min_a E\left((\theta - a)^2 \mid x\right)$. The mean of $\theta \mid x$ minimizes this, i.e. the posterior mean.

Consistency

Consistent estimators

An estimator $\delta_n(X) = \delta(X_1, \ldots, X_n)$ is consistent if

$$\delta(X) \xrightarrow{P} \theta \quad \text{as } n \to \infty$$

Under fairly general conditions and for a wide range of loss functions, the Bayes estimator is consistent

Practice Exercises

7.2	2, 4, 6
7.3	5, 10, 12, 15
7.4	5, 6, 7