03: BAYESIAN ESTIMATION

Larsen & Marx 5.8

Prof Amanda Luby

1 Bayes Theorem

Bayesian statistics is a set of techniques that are based on inverse probabilities calculated using Bayes' theorem. Relative to "classical techniques" (MoM and MLE), Bayesian estimation provides a way to incorporate "prior knowledge" into the estimation of parameters.

Example:

Classical Statistics	Bayesian Statistics
Probability refers to limiting relative frequencies. Probabilities are objective properties of the real world.	*Probability* describes a degree of belief, not a limiting frequency. As such, we can make probability statements about lots of things, not just data which are subject to random variation. For example, I might say that "the probability that Albert Einstein drank a cup of tea on August 1, 1948 is .35". This does not refer to any limiting frequency. It reflects my strength of belief that the proposition is true.
Parameters are fixed, unknown constants, and the data we observe is random. Because they are constant, no useful probability statements can be made about parameters.	*Parameters* are random, and the data that we observe are fixed. We can therefore make probability statements about parameters.
Statistical procedures should be designed to have well-defined long-run frequency properties. For example, a 95% confidence interval should capture the true value of the parameter at least 95% of the time.	We make inferences about a parameter θ by producing a probability distribution for θ. Inferences, such as point estimates and interval estimates, may then be extracted from this distribution.

Bayesian inference is a controversial approach because it inherently embraces a subjective notion of probability. The field of statistics generally puts more emphasis on frequentist methods although Bayesian methods definitely have a presence.

2 Bayesian Inference

1. **Prior distribution:**

2. **Statistical model for data:**

3. **Posterior distribution:**

Posterior distribution

Let W be a statistic dependent on parameter θ. Call its pdf $f_W(w|\theta)$. Assume that θ is the value of a random variable Θ, whose prior distribution is denoted p_Θ if discrete and f_Θ if continuous. The *posterior distribution* of Θ given $W = w$ is:

4. **Posterior mean**:

Example: Let $X_1, ..., X_n \sim \text{Bernoulli}(p)$ and suppose that p has the prior distribution $p \sim \text{Beta}(\alpha, \beta)$.

Conjugate prior

Example: Let $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$ and suppose we take $\theta \sim N(a, b^2)$. For simplicity, let's assume σ^2 is known.