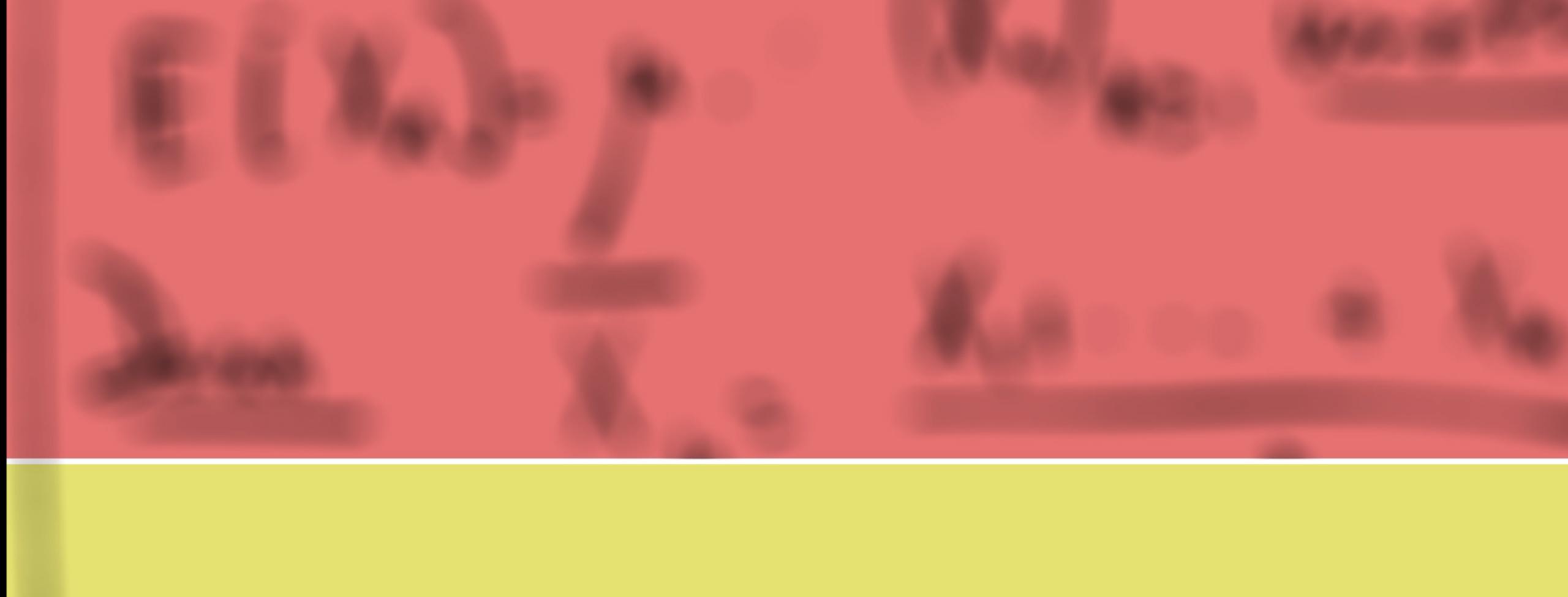
ADWANGED AND CONTRACTOR OF THE PROPERTY OF THE



BAYESIAN FUNDAMENTALS: CONCEPTS, NOTATION, AND TERMINOLOGY

Notation

$$y = random data (observed)$$

$$\theta$$
 = parameter (unobserved)

In Bayesian analysis, θ is also random. Its distribution expresses our uncertainty about its value.

Both y and θ can be vectors, e.g.,

$$y = (y_1, \dots, y_n)$$

2

All distributions we will use are defined through densities.

Reminder: We denote

$$p(\cdot)$$
 = any joint or marginal density

$$p(\cdot \mid \cdot)$$
 = any conditional density

with the symbols of the arguments specifying which random variables are involved.

Note: Discrete densities and continuous densities will use the same notation – distinguish according to the type of random variable involved.

:

Distribution Concepts

A **Bayesian model** is specified by the prior and sampling distributions:

$$p(\theta) = ext{the prior} ext{ density}$$
 $p(y \mid \theta) = ext{the sampling density}$

Together, these define any probability statement about y and θ , whether joint, marginal, or conditional.

Regard the prior as a summary of the uncertainty about θ , before observing y.

4

Bayesian inference about θ is based on

$$p(\theta \mid y)$$
 = the **posterior** density

which defines the posterior distribution.

Regard the posterior as a summary of the remaining uncertainty about θ , after observing y.

Classical non-Bayesian inference, in contrast, does not treat θ as random, so cannot summarize its uncertainty with a distribution.

ŗ

Bayes' Rule

Bayes' rule specifies how to derive the posterior from the Bayesian model:

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta) p(y \mid \theta)}{p(y)}$$

The **normalizing factor** p(y) ensures that $p(\theta \mid y)$ is a density in θ :

$$p(y) \ = \ \left\{ \begin{array}{rcl} & \displaystyle \sum_{\theta} p(\theta) \, p(y \mid \theta) & & \theta \text{ discrete} \\ & \displaystyle \int p(\theta) \, p(y \mid \theta) \, d\theta & & \theta \text{ continuous} \end{array} \right.$$

6

Since the normalizing factor is just a constant (calculated from the observed y), Bayes' rule can be written in an **unnormalized** form:

$$p(\theta \mid y) \propto p(\theta) p(y \mid \theta)$$

where the proportionality is in θ (not y).

Since the normalizing factor p(y) can be difficult to compute, we will prefer methods that do not require it.

The factor $p(y \mid \theta)$, when regarded as a function of θ for fixed y, is the **likelihood**. It is generally *not* a density in θ .