Statistics March 24, 2021

Frequentist

1. Probability is consistent with long term relative frequency.
2. Parameters are fixed unknown constants, not random quantities.
3. Statistical procedures should have well-defined long-run properties.

Bayesian

1. Probability is a degree of belief that an event will occur.
2. Can make probabilistic statements about parameters even though they are fixed constants.
3. Parameter inference is done by computing a probability distribution of θ, point estimates are computed after the fact.

Example of a Bayesian Method

Goal make a statement about some unknown parameter θ.

1. Choose a prior distribution $f(\theta)$ which captures initial belief about θ.
2. Choose a statistical model $f(\vec{x} \mid \theta)$ which captures beliefs about data $\vec{x}$ given the parameter θ: conditional distribution

Note: Write $f(\vec{x} \mid \theta)$ insted of $f(\vec{x}; \theta)$.

3. Observe $X_1, \ldots, X_n$, update our belief about θ in the form: $f(\theta \mid \vec{x})$ ← the posterior distribution

The posterior is just a conditional distribution:

$$f(\theta \mid \vec{x}) = \frac{f(\theta, \vec{x})}{f(\vec{x})} = \frac{f(\vec{x} \mid \theta)\, f(\theta)}{\int f(\theta, \vec{x})\, d\theta}$$

$$= \frac{f(\vec{x} \mid \theta)\, f(\theta)}{\int f(\vec{x} \mid \theta)\, f(\theta)\, d\theta} \quad \Big\} \text{ Bayes Theorem.}$$

In the case of n IID observations $X_1, \dots, X_n$,

$$f(\vec{x} \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta) = \mathcal{L}(\theta);$$

and therefore

$$\underbrace{f(\theta \mid \vec{x})}_{\text{function of } \theta} = \underbrace{\frac{f(\vec{x} \mid \theta)\, f(\theta)}{\int f(\vec{x} \mid \theta)\, f(\theta)\, d\theta}}_{\text{a constant } C = C(\vec{x})} = \frac{\mathcal{L}(\theta)\, f(\theta)}{C}$$

The constant C is such that $\int f(\theta \mid \vec{x})\, d\theta = 1$

$$= \int \frac{\mathcal{L}(\theta)\, f(\theta)}{C}\, d\theta.$$

Often for this reason, we will write

$$\underbrace{f(\theta \mid \vec{x})}_{\text{posterior}} \propto \underbrace{\mathcal{L}(\theta)}_{\text{model}}\; \underbrace{f(\theta)}_{\text{prior}}.$$

To generate a point estimate compute a functional of the posterior $f(\theta \mid \vec{x})$:

$$\underbrace{\bar{\Theta}}_{\text{Bayesian estimator}} = \mathbb{E}(\theta \mid \vec{x}) = \int \theta\, f(\theta \mid \vec{x})\, d\theta = \frac{\int \theta\, \mathcal{L}(\theta)\, f(\theta)\, d\theta}{\int \mathcal{L}(\theta)\, f(\theta)\, d\theta}$$

Posterior α interval (not a confidence interval)

Find a, b such that

$$P(\theta \mid \vec{x} \in (a,b)) = 1-\alpha = \int_a^b f(\theta \mid \vec{x})\, d\theta .$$

Example $X_1, \dots, X_n \sim \text{Bernoulli}(p)$ random variables

Our model : $f(\vec{x} \mid p) = \mathcal{L}(p)$

$$= \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

prior : $p \sim \text{Uniform}(0,1)$

$f(p) = 1$ on $(0,1)$.

posterior $f(p \mid \vec{x}) \propto \mathcal{L}(p)\, f(p)$

$$= \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$= p^{s} (1-p)^{n-s} \qquad s = \sum_{i=1}^{n} x_i .$$

think of this as a function of p, not the data x_i anymore

$$= p^{(s+1)-1} (1-p)^{(n-s+1)-1}$$

Now identify which family of probability distribution $\mathcal{L}(p)\, f(p)$ belongs to.

Recall: Beta(α, β) density : $f(p; \alpha, \beta) = \dfrac{\Gamma(\alpha,\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, p^{\alpha-1} (1-p)^{\beta-1}$

$$\Rightarrow f(p \mid \vec{x}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, p^{\alpha-1} (1-p)^{\beta-1} \quad \text{with} \quad \alpha = 1 + \sum x_i, \quad \beta = 1 + n - \sum x_i$$

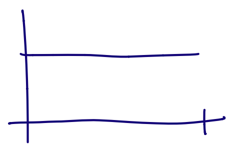
$\Rightarrow p \mid \vec{x} \sim \text{Beta}(\alpha, \beta)$.

Bayes estimator:

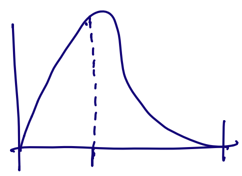
$$\bar{p} = \mathbb{E}(p \mid \vec{x}) = \underbrace{\frac{s+1}{n+2}}_{\text{mean of } \beta(\alpha,\beta) = \frac{\alpha}{\alpha+\beta}.}$$

<u>Graphically</u>

<u>Uniform prior</u>



<u>Beta</u>



<u>Part 2</u> If we use the prior $p \sim \mathrm{Beta}(\alpha, \beta)$, then repeating the calculation gives

$$p \mid \vec{x} \sim \mathrm{Beta}(\alpha + s, \beta + n - s).$$

(Note: that $p \sim \mathrm{Unifor}(0,1)$ is just $p \sim \mathrm{Beta}(1,1)$)

This is a case when the prior family <u>equals</u> the posterior family.

We say that "the prior is <u>conjugate</u> with respect to the model".

$$f(p \mid \vec{x}) \propto \overbrace{f(\vec{x} \mid p)}^{\text{model}} \; f(p)$$

same family

Functions of parameters

- Recall from MLE that if $\hat{p}$ is the MLE estimate for p, then the MLE estimate for $\tau = g(p)$ was just $\hat{\tau} = g(\hat{p})$.
- Furthermore if $Y = g(X)$, then

$$F(y) = \mathbb{P}(Y \le y) = \mathbb{P}(g(X) \le y) = \int_{g(x)\le y} f(x)\,dx$$

and $f(y) = F'(y)$.

We can use these ideas for Bayesian inference as well.

Let $\tau = g(\theta)$.

Bayes says : $f(\theta|\vec{x}) \propto \mathcal{L}(\theta)\, f(\theta)$.

Posterior CDF for $\tau = g(\theta) \,|\, \vec{x}$

$$= H(\tau|\vec{x})$$
$$= \mathbb{P}(g(\theta) \le \tau \,|\, \vec{x})$$
$$= \int_{g(\theta)\le\tau} f(\theta|\vec{x})\,d\theta$$

Then the posterior $h(\tau|\vec{x}) = \frac{d}{d\tau} H(\tau|\vec{x})$.

<u>Example</u> $X_i \sim \text{Bernoulli}(p)$, prior $f(p) = 1$.

$\Rightarrow$ $p|\vec{x} \sim \text{Beta}(s+1, n-s+1)$. Let $\psi = \log\left(\frac{p}{1-p}\right)$.

$p \in (0,1) \Rightarrow \psi \in (-\infty, \infty)$.

$$\Rightarrow H(\psi|\vec{x}) = \mathbb{P}\left(\log\frac{p}{1-p} \le \psi \,\middle|\, \vec{x}\right)$$
$$= \mathbb{P}\left(p \le \frac{e^\psi}{1+e^\psi} \,\middle|\, \vec{x}\right)$$

$$= \int_0^{e^\psi/1+e^\psi} f(p \mid \vec{x})\, dp$$

$$= \int_0^{e^\psi/1+e^\psi} \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} p^s (1-p)^{n-s}\, dp$$

To compute $h(\psi \mid \vec{x})$, compute $\frac{d}{d\psi} H(\psi \mid \vec{x})$:

$$h(\psi \mid \vec{x}) = \frac{d}{d\psi} H(\psi \mid \vec{x})$$

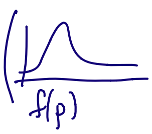
$$= \frac{d}{d\psi}\left(\int_0^{e^\psi/1+e^\psi} \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} p^s (1-p)^{n-s}\, dp \right)$$

$$= \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} e^\psi \left(\frac{e^\psi}{1+e^\psi}\right)^s \left(\frac{1}{1+e^\psi}\right)^{n-s}$$

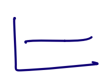
Types of Priors

The big question in Bayesian analysis is where does the prior come from in a general problem.

Subjective prior vs. **Non-informative prior**

e.g. (

, normal distribution)

e.g. ( , flat, uniform pdf).

Improper prior: Let $x \mid \mu \sim N(\mu, \sigma^2)$, σ^2 is known.

Set a flat prior on μ: $f(\mu) = c > 0$.

$$\Rightarrow \int_{-\infty}^{\infty} f(\mu)\, d\mu = \infty \quad \leftarrow \text{not a prob. density.}$$

This is known as an improper prior, but Bayes can formally be carried out:

$$f(\mu|x) = \frac{\mathcal{L}(\mu)\cdot f(\mu)}{\int \mathcal{L}(\mu)\cdot f(\mu)\, d\mu}$$

$$= \frac{\mathcal{L}(\mu)\cdot \cancel{c}}{\int \mathcal{L}(\mu)\cdot \cancel{c}\, d\mu} = \frac{\mathcal{L}(\mu)}{\underbrace{\int \mathcal{L}(\mu)\, d\mu}}$$

$$\hookrightarrow \int_{-\infty}^{\infty} \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}\, d\mu = 1$$

$$\Rightarrow \mu|x \sim N(x, \sigma^2)$$

In general, improper priors are not a problem so long as $f(\vec{x}|\theta)$ decays fast enough as a function of θ.

Flat priors are not transformation invariant

Logically, if we know nothing about a parameter p, the we should also know nothing about $\psi = \log\left(\frac{p}{1-p}\right)$, but $f(p)=1 \Rightarrow f(\psi) \neq 1$. Contradiction?

My interpretation: flat priors do not mean non-informative.

Jeffrey's Prior

Set $f(\theta) \propto \sqrt{I(\theta)}$ ← Fisher information

The Jeffrey's Prior is transformation invariant.

If $\tau = g(\theta)$, then what does the prior for τ look like?

$$I(\theta) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log f(\vec{x};\theta) \right)$$

statistical model.

$$= -\mathbb{E}\left(\left(\frac{d\tau}{d\theta}\right)^2 \frac{\partial^2}{\partial \tau^2} \log f(\vec{x};\tau) \right)$$

$$= -\left(\frac{d\tau}{d\theta}\right)^2 \mathbb{E}\left(\frac{\partial^2}{\partial \tau^2} \log f(\vec{x};\tau) \right)$$

$$\doteq \left(\frac{d\tau}{d\theta}\right)^2 I(\tau)$$

$$\Rightarrow \sqrt{I(\theta)} = \sqrt{I(\tau)} \left| \frac{d\tau}{d\theta} \right|$$

This the usual change of variable formula.