- Lecture 6

Posterior predictive distribution – finding X_4 from $X_1 X_2 X_3$

$\theta \sim P(\theta)$ – prior

$X_1 X_2 X_3$ – data observed

We want to find:

$X_1 X_2 X_3 X_4 \overset{iid}{\sim} Bern(\theta)$

Note: $P(Y) = \sum_{x \in supp(X)} P(X, Y)$ so:

- $P(X_4 \mid X_1 X_2 X_3) = \int_{\Theta_0} P(X_4 \theta \mid X_1 X_2 X_3)$

or $\sum_{\theta \in \Theta_0} P(X_4 \theta \mid X_1 X_2 X_3)$

Note: $P(X \theta) = P(X \mid \theta) P(\theta)$ so:

- $\sum_{\theta \in \Theta_0} P(X_4 \theta \mid X_1 X_2 X_3) = \sum_{\theta \in \Theta_0} P(X_4 \mid \theta X_1 X_2 X_3)^{*} P(\theta \mid X_1 X_2 X_3)$

*Note: $P(X_4 \mid \theta X_1 X_2 X_3) = P(X_4 \mid \theta)$

$= \frac{P(X_4 \theta X_1 X_2 X_3)}{P(\theta X_1 X_2 X_3)} = \frac{P(X_4 X_1 X_2 X_3 \mid \theta) P(\theta)}{P(X_1 X_2 X_3 \mid \theta) P(\theta)} \overset{(\text{since iid})}{=} \frac{P(X_4 \mid \theta) P(X_1 \mid \theta) P(X_2 \mid \theta) P(X_3 \mid \theta)}{P(X_1 \mid \theta) P(X_2 \mid \theta) P(X_3 \mid \theta)}$

so:

- $\boxed{P(X_4 \mid X_1 X_2 X_3) = \sum_{\theta \in \Theta_0} P(X_4 \mid \theta) P(\theta \mid X_1 X_2 X_3)}$ (posterior, prior)

Note: $\neq P(X_4 \mid \hat{\theta}_{MLE})$

- Having θ range over all $\Theta = (0,1)$

X_1, X_2, X_3 observed

example $\Theta_0 = \{0.25, 0.75\}$ $\quad X = \langle 0,1,1 \rangle$

$$P(\theta) = \underset{\text{uniform}}{\cup} P(\Theta_0) = \begin{cases} 0.25 \text{ w.p } \frac{1}{2} \\ 0.75 \text{ w.p } \frac{1}{2} \end{cases}$$

- 1st Goal: Point estimation (improved)

$$\hat{\theta}_{MAP} := \underset{\theta \in \Theta_0}{\arg\max} \{P(\theta|X)\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \left\{ \frac{P(X|\theta)P(\theta)}{P(X)} \right\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{P(X|\theta)\} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \{f(\theta; X)\} = \hat{\theta}_{MLE}$$

($P(\theta)$: constant if prior is uniform; $P(X)$: constant in θ)

* This ignores possible value for θ in the space

- for ex: it was shown $\hat{\theta}_{MAP} = 0.75 \neq \hat{\theta}_{MLE} = 0.66$. B/C $\Theta_0 \neq$ [illegible]*

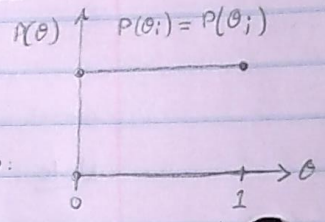
We want:

* • Parameter space of the likelihood model = support of the prior

$$[\Theta = (0,1) = \text{supp}[\theta] = (0,1)]^*$$

* uniform, so:

$$P(\theta) = \begin{cases} 1 \text{ if } \theta \in (0,1) \\ 0 \text{ otherwise} \end{cases}$$



- for ex: we want given X_1, X_2, X_3: $P(\theta)$: $\theta \sim U(0,1)$ ** standard uniform r.v

$$\hat{\theta}_{MAP} = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \left\{ \frac{P(X|\theta)P(\theta)}{P(X)} \right\} = \operatorname{argmax} \{P(X|\theta)\} = \underset{\theta \in (0,1)}{\operatorname{argmax}} \{\theta^2(1-\theta)\}$$

($P(\theta)$: constant = 1; $P(X)$: constant for $\theta \in \Theta_0$; θ^2: $x=1$; $(1-\theta)$: $x=0$)

* Note to find max set $\frac{d}{d\theta}[P(X|\theta)] = 0$

- so: $\frac{d}{d\theta}[\theta^2(1-\theta)] = \frac{d}{d\theta}[\theta^2 - \theta^3] = 2\theta - 3\theta^2 = 0 \Rightarrow 2\theta = 3\theta^2 \Rightarrow 2 = 3\theta$

$\theta = \frac{2}{3}$

Now $\hat{\theta}_{MAP} = \frac{2}{3} = \hat{\theta}_{MLE}$

* frequentism this isnt possible, only 1 (in range) or 0 (not in range)

- 2nd Goal: Probability of falling in some range.* $P(\theta \in [0.6, 0.7] \mid X = \langle 0,1,1 \rangle)$

$$\int_{0.6}^{0.7} P(\theta|X)\,d\theta = \int_{0.6}^{0.7} \frac{P(X|\theta)P(\theta)}{P(X)}\,d\theta = \int_{0.6}^{0.7} \frac{\theta^2(1-\theta)\cdot(1)}{\int_\Theta P(X|\theta)P(\theta)\,d\theta}\,d\theta$$

$$= \int_{0.6}^{0.7} \frac{\theta^2 - \theta^3}{\frac{1}{12}}\,d\theta = 12\left[\frac{\theta^3}{3} - \frac{\theta^4}{4}\right]_{0.6}^{0.7} \approx 17.65\%$$

$$= \int_0^1 \theta^2(1-\theta)\cdot 1\,d\theta = \frac{\theta^3}{3} - \frac{\theta^4}{4}\Big|_0^1 = \left(\frac{1}{3} - \frac{1}{4}\right) - 0 = \frac{1}{12}$$

Chance θ is between $[0.6, 0.7]$

Generalization: Arbitrary data set, n realizations of the iid Bernoulli model: $X_1, X_2 \dots X_n$ / $\theta^x (1-\theta)^{1-x}$

for single X: $P(\theta|X) = \frac{P(X|\theta)\,P(\theta)}{P(X)} = \frac{P(X|\theta)\cdot 1}{\int_\Theta P(X|\theta)\,P(\theta)\,d\theta} = \frac{P(X|\theta)}{\int_\Theta P(X|\theta)\,d\theta}$

for X_i, $\{i=1,2,\dots,n\}$: $P(\theta|X_1 X_2 \dots X_n) = \frac{\prod_{i=1}^{n} P(X_i|\theta)}{\int_0^1 \prod_{i=1}^{n} P(X_i|\theta)} = \frac{\theta^{\Sigma X_i}(1-\theta)^{n-\Sigma X_i}}{\int_0^1 \theta^{\Sigma X_i}(1-\theta)^{n-\Sigma X_i}\,d\theta}$

- **Note** where X_i is either 1 or 0, so ΣX_i is all 'successes' and $n-\Sigma X_i$ is 'failures'

$= \frac{\theta^{\Sigma X_i}(1-\theta)^{n-\Sigma X_i}}{\int_0^1 \theta^{(\Sigma X_i+1)-1}(1-\theta)^{(n-\Sigma X_i+1)-1}\,d\theta}$ } Beta function

- **Note** Beta functions are: $B(\alpha,\beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$

$P(\theta|X) \left[= \frac{\theta^{(\Sigma X_i+1)-1}(1-\theta)^{(n-\Sigma X_i+1)-1}}{B(\Sigma X_i+1,\ n-\Sigma X_i+1)} \right] := \left[\text{Beta}\left(\Sigma X_i+1,\ n-\Sigma X_i+1\right) \right]$

PDF for Beta distribution

- Properties of the Beta distribution

Let $Y \sim \text{Beta}(\alpha,\beta) := \frac{1}{B(\alpha,\beta)} \cdot y^{\alpha-1}(1-y)^{\beta-1} = f(y)$

- $\text{supp}[Y] = (0,1)$
- parametric space: $\alpha > 0$, $\beta > 0$

is $f(y)$ a PDF? does $\int_{\text{supp}[Y]} f(y)\,dy = 1$?

$\int_{\text{supp}(y)} f(y)\,dy = \int_0^1 \frac{1}{B(\alpha,\beta)} \cdot y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{1}{B(\alpha,\beta)} \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy = \frac{B(\alpha,\beta)}{B(\alpha,\beta)} = 1$

so $f(y)$ is a PDF

- **Note:** Gamma function

$\Gamma(\alpha) := \int_0^\infty t^{\alpha-1} e^{-t}\,dt$ for $\forall\, \alpha > 0$

properties

ⓐ $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$ (integration by parts)

ⓑ $B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ (mult. var. calc.)

- Note α = success + 1
  β = n − (success + 1)

$Y \sim Beta(\alpha, \beta)$

- Expectation

$$E[Y] = \int_{supp[Y]} y\, f(y) = \int \frac{y \cdot y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} = \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \Big/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

use properties of gamma function

$$\alpha\Gamma(\alpha) \quad \alpha+\beta\,\Gamma(\alpha+\beta)$$

$$\boxed{E[Y] = \frac{\alpha}{\alpha+\beta}}$$

- Mode

$$Mode[Y] = \underset{y \in supp[Y]}{argmax}\ \{f(y)\} = \underset{y\in(0,1)}{argmax}\ \left\{\frac{1}{B(\alpha,\beta)} \cdot y^{\alpha-1}(1-y)^{\beta-1}\right\} = \underset{y\in(0,1)}{argmax}\ \{y^{\alpha-1}(1-y)^{\beta-1}\}$$

(greatest Y value)

↑ constant

$$= \ell \underset{y\in(0,1)}{argmax}\ \{(\alpha-1)\ln(y) + (\beta-1)\ln(1-y)\}$$

- Note: find argmax by $\frac{d}{dy}[\ell(y)] = 0$

$$\Rightarrow \frac{d}{dy}[\ \|\ \|\ ] = \frac{\alpha-1}{y} - \frac{\beta-1}{1-y} = 0$$

$$\Rightarrow \frac{\alpha-1}{y} = \frac{\beta-1}{1-y} \Rightarrow \boxed{y = \frac{\alpha-1}{\alpha+\beta-2} \quad \text{when } \alpha>1,\ \beta>1} = Mode[Y]$$

- Beta r.v. density graphed: