Recall the example where $X|\theta \sim \text{Binomial}(n,\theta)$ and we observe

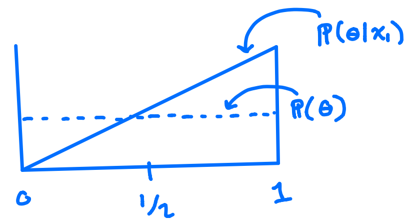
$n=12,\ X=9.$

For now, assume we observe the sequence of outcomes

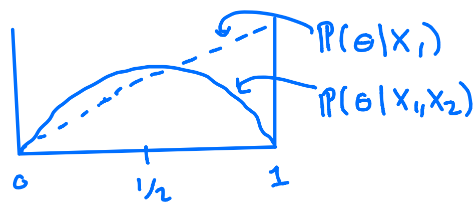
$$(1,0,1,1,0,1,\dots,1) \overset{\text{def}}{=} (x_1,x_2,\dots,x_{12})$$

Let's assume $\mathbb{P}(\theta) \propto 1 \implies \theta \sim \text{Unif}(0,1)$

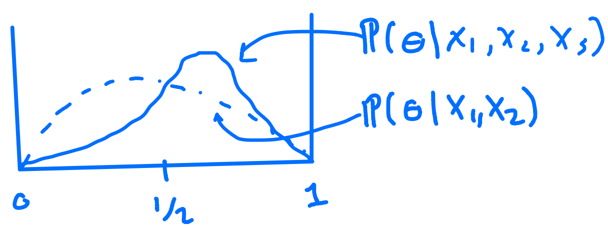
1. $\mathbb{P}(\theta|x_1) \propto \mathbb{P}(x_1|\theta)\,\mathbb{P}(\theta) = \theta$



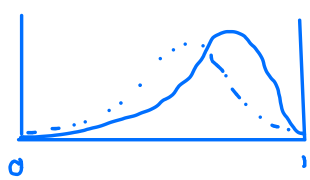
2. $\mathbb{P}(\theta|x_1,x_2) \propto \mathbb{P}(x_2|\theta)\,\mathbb{P}(\theta|x_1) = (1-\theta)\theta$



3. $\mathbb{P}(\theta|x_1,x_2,x_3) \propto \mathbb{P}(x_3|\theta)\,\mathbb{P}(\theta|x_1,x_2) = \theta^2(1-\theta)$

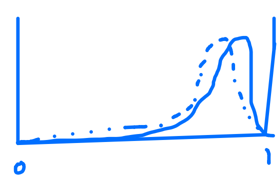


9.



. . .

12.



Rmks: As data size increases, variance decreases

$$\mathbb{P}(\theta|X) \propto \mathbb{P}(X_{12}|\theta)\mathbb{P}(\theta|X_1,\dots,X_{11}) = \theta^9(1-\theta)^{12-9}$$

highlights: (a) sufficient statistics (b) order doesn't matter / exchangeable

Rmk: Practically, $\mathbb{P}(\theta|X)$ is all that matters. Still needs to summarize the posterior.

Summarize Posterior for inferences.

ex: point estimate: $\hat{\theta} = \mathbb{E}[\theta|X]$ has the property

$$\hat{\theta} = \underset{\tilde{\theta}}{\operatorname{argmin}}\ \mathbb{E}_{\theta|X}[(\theta-\tilde{\theta})^2]$$

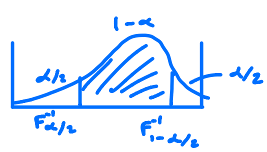
ex: Interval estimators — credible intervals

$$\mathbb{P}(\theta \in S_\alpha | X) = 1-\alpha$$

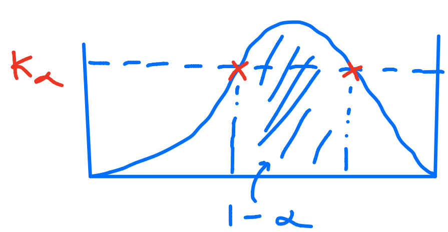
random quant.

Constructing Credible Intervals

(i) Percentage Cred. interval: if F is post. CDF then $\mathbb{P}\left(\theta \in \left(F^{-1}_{\alpha/2}, F^{-1}_{1-\alpha/2}\right)\right) = 1-\alpha$



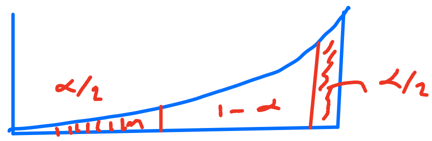
(ii) Highest Post. density (HPD) Credible Intervals



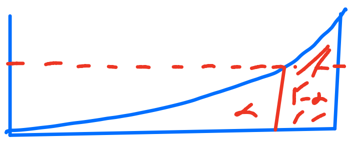
$\{\theta : \mathbb{P}(\theta|x) \geq k_\alpha\}$ with k_α s.t.

$$\mathbb{P}(\{\theta : \mathbb{P}(\theta|x) \geq k_\alpha\}|x) = 1-\alpha$$

Ex:

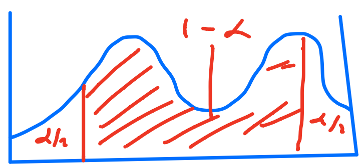


Percentage

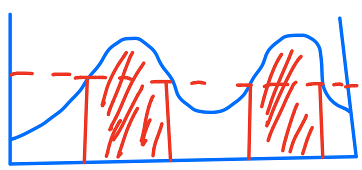


HPD

Rmk: HPD guarteend to have point of most mass.



Percentage – includes low density region



HPD – disconnected interval

Going back to the example.

$X|\theta \sim Bin(n,\theta)$; $\mathbb{P}(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \Rightarrow \mathbb{P}(\theta) = \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$

where $B(\alpha,\beta)$ is s.t. $\int_0^1 \mathbb{P}(\theta)d(\theta) = \frac{1}{B(\alpha,\beta)}\int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = 1$

$\Rightarrow B(\alpha,\beta) = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta$ "Beta function"

Ex: $\theta \sim Unif(0,1) \overset{D}{=} Beta(1,1)$

Properties: mode $= \frac{\alpha-1}{\alpha+\beta-2}$, $\mathbb{E}[\theta] = \frac{\alpha}{\alpha+\beta}$, $Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$= \frac{\mathbb{E}[\theta](1-\mathbb{E}[\theta])}{\alpha+\beta+1}$ Dispersed bernoulli variance

__Posterior:__ $$\mathbb{P}(\theta|X) = \frac{\mathbb{P}(X|\theta)\,\mathbb{P}(\theta)}{\int_0^1 \mathbb{P}(X|\theta)\,\mathbb{P}(\theta)\,d\theta} = \frac{\cancel{\binom{n}{x}}\,\theta^x(1-\theta)^{n-x}\,\cancel{\frac{1}{B(\alpha,\beta)}}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 \cancel{\binom{n}{x}}\,\theta^x(1-\theta)^{n-x}\,\cancel{\frac{1}{B(\alpha,\beta)}}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}\,d\theta}$$

$$= \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\,d\theta} = \frac{1}{B(x+\alpha,n-x+\beta)}\,\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

$$= \mathrm{Beta}(x+\alpha,\, n-x+\beta)$$

Therefore, $\theta|X \sim \mathrm{Beta}(x+\alpha,\, n-x+\beta)$

__posterior mode__ is then $\frac{x+\alpha-1}{x+\alpha+n-x+\beta} = \frac{x+\alpha-1}{n+\alpha+\beta-2}$

if uniformative, $\alpha=\beta=1$ $\quad \frac{x+1-1}{n+1+1-2} = \frac{x}{n} =$ ML estimator $= \hat{\theta}_{ML}$

__Posterior Mean__ $$\frac{x+\alpha}{n+\alpha+\beta} = \underbrace{\frac{x}{n}}_{\hat{\theta}_{ML}}\frac{n}{n+\alpha+\beta} + \underbrace{\frac{\alpha}{\alpha+\beta}}_{\hat{\theta}_{\text{prior mean}}}\cdot\frac{\alpha+\beta}{n+\alpha+\beta}$$

$$\frac{n}{n+\alpha+\beta} \xrightarrow{n} 1, \quad \frac{\alpha+\beta}{n+\alpha+\beta} \xrightarrow{n} 0$$

If $\alpha=\beta=1$, $\mathbb{E}[\theta|X] = \frac{x+1}{n+2}$ "Laplace's law of succession"

__Posterior Variance__

$$\mathrm{Var}(\theta|X) = \frac{\mathbb{E}[\theta|X](1-\mathbb{E}[\theta|X])}{n+\alpha+\beta+1} \xrightarrow{n} 0$$

__Def:__ Bayesian consistency: $\hat{\theta}_{post} \xrightarrow{\mathbb{P}} \hat{\theta}_{ML}$