1/11/21

① remote instruction for two more weeks (until 1/31)

② office hours — remote
available for in-person office hours
when in-person instruction becomes ok.

$B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

For given $\theta \in (0,1)$

$x \mid \theta \sim \text{Binomial}(n, \theta)$, $x \in \{0, 1, \dots, n\}$

PF Sec 3.1 : Binomial model

† PF §1.2.1 Example (contd): Suppose that ~~the sampling model for x_i is a Bernoulli, i.e., $x_i \mid \theta \overset{iid}{\sim}$ Ber(θ)~~, and the prior is Be(α, β), where the hyperparameters α and β are known,

$\theta \in (0,1)$

$\alpha > 0$, $\beta > 0$

$$\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < \theta < 1.$$

Find the joint, marginal, posterior, and predictive distributions.

$\theta \sim Be(\alpha, \beta)$ $\qquad E(\theta) = \dfrac{\alpha}{\alpha+\beta}$ $\qquad \alpha \uparrow, \quad E(\theta) \uparrow$

$Var(\theta) = \dfrac{\alpha\beta}{\underline{(\alpha+\beta)^2} \cdot (\alpha+\beta+1)}$ $\qquad \alpha+\beta \uparrow, \quad Var(\theta) \downarrow$

$\alpha+\beta$: prior sample size

α : # of prior successes

β : # of prior failures

① $h(\theta, x)$

$$h(\theta, x) = f(x \mid \theta)\, \pi(\theta)$$

$$= \binom{n}{x} \theta^{x} (1-\theta)^{n-x} \cdot \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$= \binom{n}{x} \frac{1}{B(\alpha,\beta)} \theta^{\alpha+x-1} (1-\theta)^{n-x+\beta-1}, \quad x \in \{0, \dots, n\},\ \theta \in (0,1)$$

② $m(x)$

$$m(x) = \int_0^1 h(\theta, x)\, d\theta$$

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

$$= \binom{n}{x} \frac{1}{B(\alpha,\beta)} \underbrace{\int_0^1 \theta^{\alpha+x-1} (1-\theta)^{n-x+\beta-1}\, d\theta}_{\text{a kernel for } Be(\alpha+x,\ n-x+\beta)}$$

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

$$= \binom{n}{x} \frac{B(\alpha+x, n-x+\beta)}{B(\alpha, \beta)} \qquad x \in \{0, 1, \dots, n\}$$

↖ Beta-binomial distribution

③ $\pi(\theta \mid x)$

$$\pi(\theta \mid x) = \frac{h(\theta, x)}{m(x)} = \frac{f(x \mid \theta)\,\pi(\theta)}{m(x)} \propto f(x \mid \theta)\,\pi(\theta)$$

$$\pi(\theta \mid x) = \frac{\cancel{\binom{n}{x}} \frac{1}{\cancel{B(\alpha,\beta)}}\ \theta^{\alpha+x-1}\ (1-\theta)^{n-x+\beta-1}}{\cancel{\binom{n}{x}} \frac{B(\alpha+x,\, n-x+\beta)}{\cancel{B(\alpha,\beta)}}}$$

$$= \frac{1}{Be(\alpha+x,\ n-x+\beta)}\ \theta^{\alpha+x-1}\ (1-\theta)^{n-x+\beta-1}, \qquad \theta \in (0, 1)$$

$$\Rightarrow \quad \theta \mid x \sim Be(\alpha+x,\ \beta+n-x)$$

$$E(\theta) = \frac{\alpha}{\alpha+\beta}$$

$$E(\theta \mid X) = \frac{\alpha+x}{\alpha+x+\beta+n-x} = \frac{\alpha+x}{\alpha+\beta+n}$$

$$= \underbrace{\left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)}\cdot \underbrace{\frac{\alpha}{\alpha+\beta}}_{=E(\theta)} + \underbrace{\left(\frac{n}{\alpha+\beta+n}\right)}\cdot \underbrace{\left(\frac{x}{n}\right)}_{\text{sample proportion}}$$

fix α, β, $\quad n \uparrow \rightarrow \frac{n}{\alpha+\beta+n} \uparrow \rightarrow E(\theta \mid x) \approx \frac{x}{n}$

fix n, $\quad \alpha+\beta \uparrow \rightarrow \underbrace{\frac{\alpha+\beta}{\alpha+\beta+n}} \uparrow \rightarrow E(\theta \mid x) \approx E(\theta)$

④ $f(y|x)$ $\qquad$ $y|\theta \sim \text{Binomial}(m, \theta)$

x & y are cond. indep. given θ

$$f(y|x) = \int_0^1 f(y, \theta | x)\, d\theta$$

$$= \int_0^1 f(y|\theta, x)\, \pi(\theta|x)\, d\theta$$

$$= \int_0^1 \underbrace{f(y|\theta)}\, \underbrace{\pi(\theta|x)}\, d\theta$$

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

$$= \binom{m}{y} \cdot \frac{B(\alpha+x+y, \beta+n-x+m-y)}{B(\alpha+x, \beta+n-x)}, \quad y \in \{0, \dots, m\}$$

Beta-Binomial distribution.

③'

$$\pi(\theta|x) \propto f(x|\theta)\, \pi(\theta)$$

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

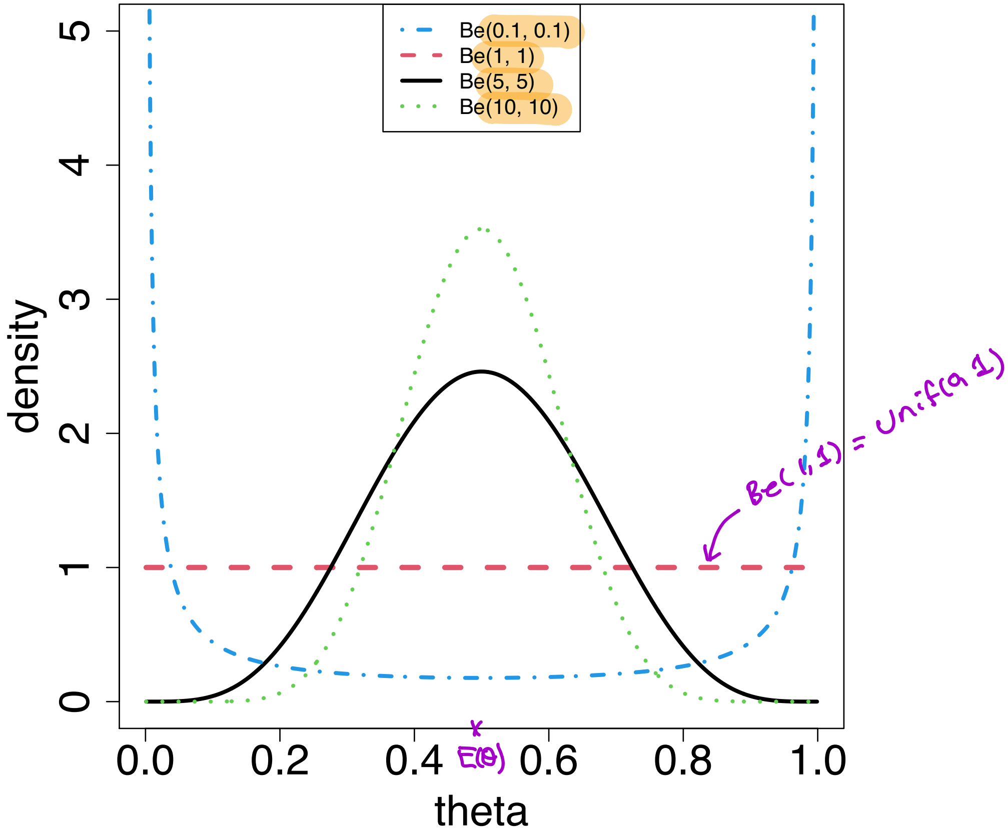
$$\propto \underbrace{\theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}}$$

→ a kernel for $Be(\alpha+x, \beta+n-x)$

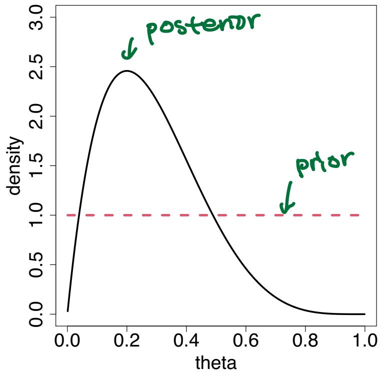
$\Rightarrow \theta|x \sim Be(\alpha+x, \beta+n-x)$

† PF §1.2.1 Example (contd): Examples of the beta density

$E(\theta) = \frac{\alpha}{\alpha+\beta} = \frac{1}{2}$



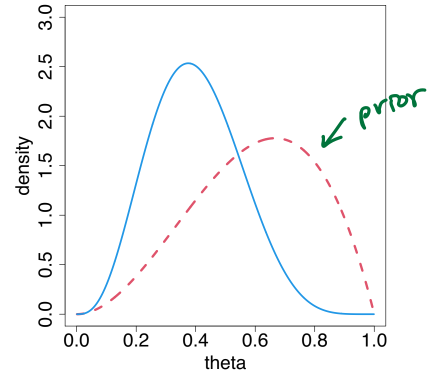
† PF §1.2.1 Example (contd):



$\alpha+\beta=2$

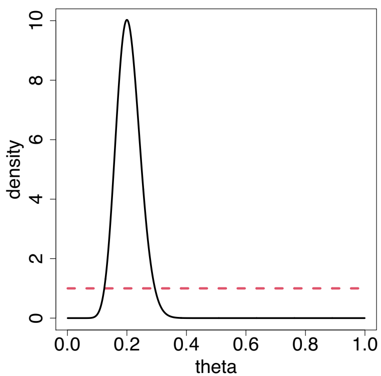
(a) Be(1, 1), $n = 5, x = 1$

$\frac{x}{n} = \frac{1}{5} = 0.2$

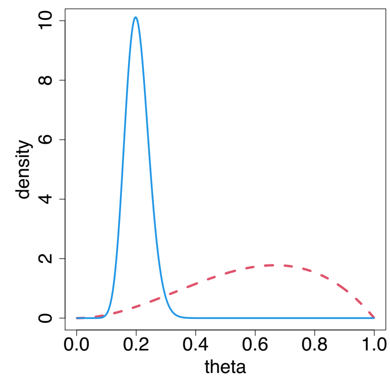


$\alpha+\beta=5$

(b) Be(3, 2), $n = 5, x = 1$

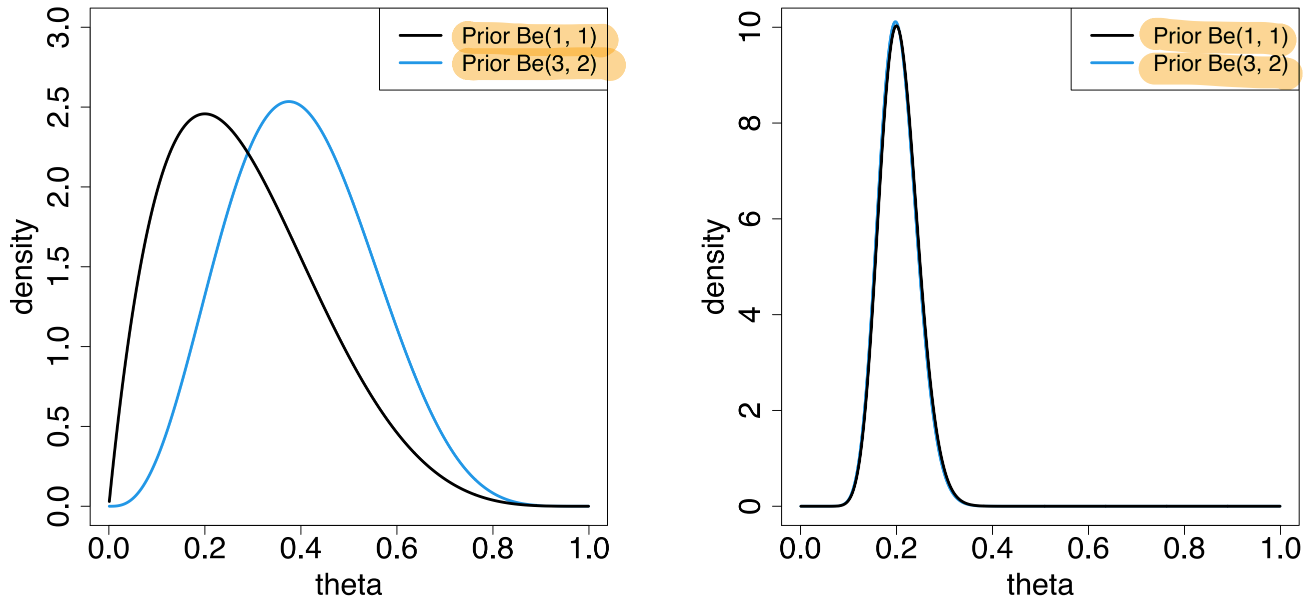


(c) Be(1, 1), $n = 100, x = 20$



(d) Be(3, 2), $n = 100, x = 20$

† PF §1.2.1 Example (contd):



(a) $n = 5, x = 1$ (b) $n = 100, x = 20$

⊕ For more, read Hoff §3.1 The binomial model.

♣ Example 2 (JB Example 1 p127): Assume that observations x_i's are normally distributed with mean θ and known variance σ^2. The parameter of interest, θ also has normal distribution with parameters μ and τ^2. Find the posterior distribution of θ given $\boldsymbol{x}$. Also, find the posterior predictive distribution of a future observation y assuming conditional independence between y and $\boldsymbol{x}$ given θ.

$**$ *Note*: This example is important because the normal likelihood and normal prior combination is very common.

$**$ Also, read Hoff §5.1-5.2.

$X_i \mid \theta \overset{iid}{\sim} N(\theta, \sigma^2)$, $i=1,\dots,n$ $\quad \sigma^2$ known

$\theta \sim N(\mu, \tau^2)$ $\quad \mu, \tau^2$: specified.

$$f(x_1,\dots,x_n \mid \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x_i-\theta)^2}{2\sigma^2}\right)$$

$(x_i \pm \bar{x} - \theta)^2$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \cdot \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\theta)^2}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{n} \frac{\left((x_i-\bar{x}) + (\bar{x}-\theta)\right)^2}{2\sigma^2}\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2\sigma^2} - \sum_{i=1}^{n} \frac{(\bar{x}-\theta)^2}{2\sigma^2} - 2\cdot\sum_{i=1}^{n} \frac{(x_i-\bar{x})\cdot(\bar{x}-\theta)}{2\sigma^2}\right\}$$

$$\frac{2\cdot(\bar{x}-\theta)}{2\sigma^2} \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{=0}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\}$$

$$\pi(\theta) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$\pi(\theta \mid \mathbf{x}) \propto f(\mathbf{x} \mid \theta)\, \pi(\theta)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\}$$

$$\times \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{(\theta-\mu)^2}{2\tau^2}\right\}$$

$$\propto \exp\left\{ -\underbrace{\frac{n(\bar{x}-\theta)^2}{2\sigma^2}} - \underbrace{\frac{(\theta-\mu)^2}{2\tau^2}} \right\}$$

$$\propto \exp\left[-\frac{1}{2} \left\{ \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)\theta^2 - 2\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right)\theta \right\} \right]$$

$$\propto \exp\left[-\frac{1}{2} \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right) \left\{ \theta^2 - 2 \cdot \underbrace{\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right)} \theta \right\} \right]$$

$$\Rightarrow \quad \theta \mid \mathbf{x} \sim N\left(\underbrace{\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right)}_{=\mu_1}, \underbrace{\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}}_{=\tau_1^2} \right)$$

$$E(\theta) = \mu$$

$$E(\theta \mid \mathbf{x}) = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right)$$

$$= \frac{\frac{1}{\sigma^2/n}\bar{x} + \frac{1}{\tau^2}\mu}{1/(\sigma^2/n) + 1/\tau^2}$$

For fixed $\bar{x}, n, \sigma^2$

$$\tau^2 \downarrow \;\to\; \frac{1}{\tau^2} \uparrow \;\to\; E(\theta \mid \mathbf{x}) \approx \mu$$

prior precision

$$n \uparrow \;\to\; 1/(\sigma^2/n) \uparrow \;\to\; E(\theta \mid \mathbf{x}) \approx \bar{x}$$

precision of $\bar{x}$

$Z \sim N(m, v^2) \Rightarrow f(z) = \frac{1}{\sqrt{2\pi v^2}} \exp\left(-\frac{1}{2v^2}(z-m)^2\right)$

$\propto \exp\left(-\frac{1}{2v^2}\left(z^2 - 2mz\right)\right)$

Assume cond. indep between ~~of~~ y and x given θ

$f(y|x) = \int_{-\infty}^{\infty} f(y|\theta, x)\, \pi(\theta|x)\, d\theta$

$= \int_{-\infty}^{\infty} \underbrace{f(y|\theta)}_{= N(\theta, \sigma^2)} \underbrace{\pi(\theta|x)}_{= N(\mu_1, \tau_1^2)} \underline{d\theta}$

$\rightarrow \quad y|x \sim N(\mu_1, \underline{\tau_1^2 + \sigma^2})$

$y|\theta \sim N(\theta, \sigma^2)$

Full derivation of the posterior predictive distribution

$$f(y \mid x) = \int_{-\infty}^{\infty} f(y \mid \theta, x)\, \pi(\theta \mid x)\, d\theta$$

$$= \int_{-\infty}^{\infty} \underbrace{f(y \mid \theta)}_{= N(\theta, \sigma^2)}\ \underbrace{\pi(\theta \mid x)}_{= N(\mu_1, \tau_1^2)}\ \underline{d\theta}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\theta)^2}{2\sigma^2}\right) \cdot \frac{1}{\sqrt{2\pi\tau_1^2}} \exp\left(-\frac{(\theta-\mu_1)^2}{2\tau_1^2}\right) d\theta$$

drop factors not having y and θ

$$\propto \exp\left(-\frac{y^2}{2\sigma^2}\right) \cdot \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left\{\left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)\theta^2 - 2\cdot\left(\frac{y}{\sigma^2}+\frac{\mu_1}{\tau_1^2}\right)\theta\right\}\right] d\theta$$

① recognize a kernel to $N\left(\left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)^{-1}\left(\frac{y}{\sigma^2}+\frac{\mu_1}{\tau_1}\right),\ \left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)^{-1}\right)$

② multiply needed factors to complete the identified density function

the terms from completing the density function

$$\propto \exp\left(-\frac{y^2}{2\sigma^2}\right) \cdot \sqrt{2\pi\left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)^{-1}} \cdot \exp\left(\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)^{-1}\left(\frac{y}{\sigma^2}+\frac{\mu_1}{\tau_1}\right)^2\right)$$

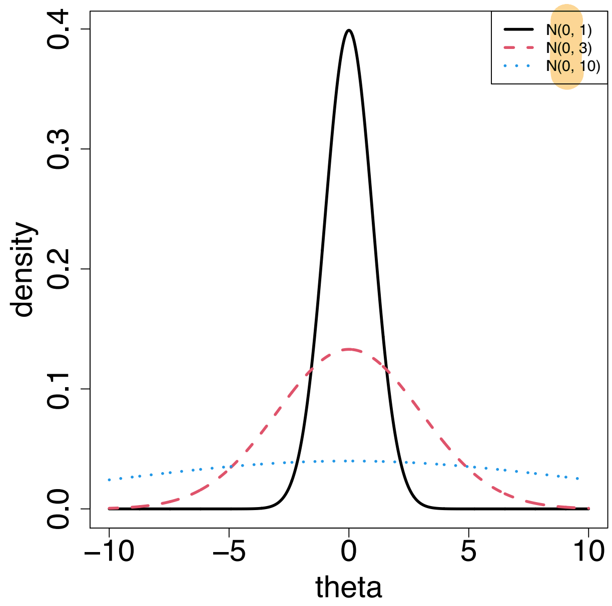
$$\propto \exp\left[-\frac{1}{2}\left\{\underbrace{\left(\frac{1}{\sigma^2} - \frac{1}{\sigma^4}\left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)^{-1}\right)}_{= \frac{1}{\sigma^2} - \frac{\tau_1^2}{\sigma^2(\sigma^2+\tau_1^2)} = \frac{1}{\sigma^2+\tau_1^2}} y^2 - 2\underbrace{\left(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}\right)^{-1}\left(\frac{\mu_1}{\tau_1\sigma^2}\right)}_{= \frac{\mu_1}{\sigma^2+\tau_1^2}} y\right\}\right]$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2+\tau_1^2}\right)\left(y^2 - 2\mu_1 y\right)\right\}$$

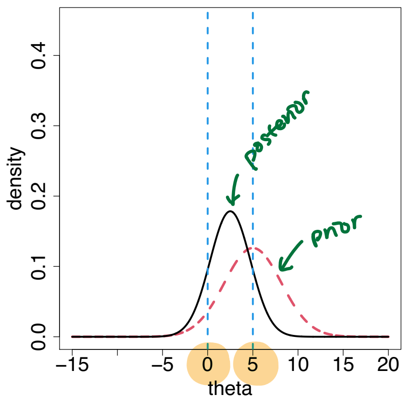
recognize a kernel to $N(\mu_1, \sigma^2 + \tau_1^2)$

$\Rightarrow \quad y \mid x \sim N(\mu_1, \sigma^2 + \tau_1^2)$

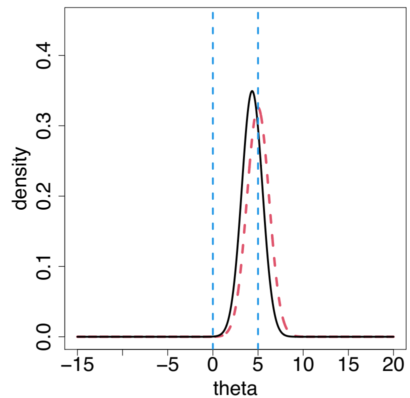
♣ Example 2(contd) Examples of the normal density



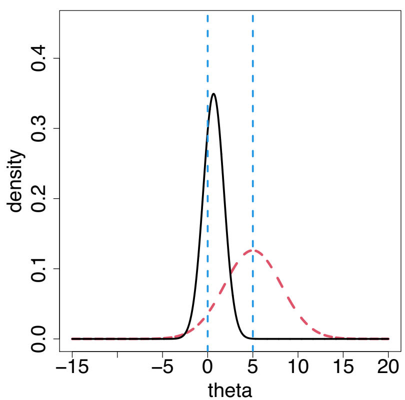
♣ Example 2(contd): Suppose $\bar{x} = 0$ with $n = 1$ and $\mu = 5$.



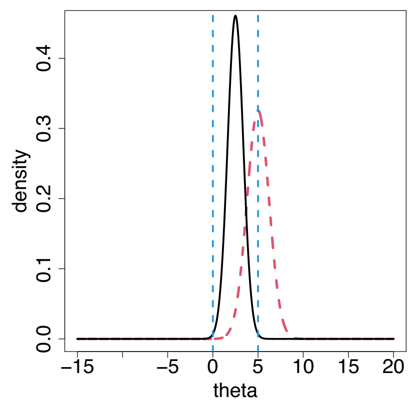
(a) $\sigma^2 = 10$ & $\tau^2 = 10$



(b) $\sigma^2 = 10$ & $\tau^2 = 1.5$



(c) $\sigma^2 = 1.5$ & $\tau^2 = 10$



(d) $\sigma^2 = 1.5$ & $\tau^2 = 1.5$

♣ Example 2(contd): Suppose $\bar{x} = 0$ and $\mu = 5$ with $\sigma^2 = \tau^2 = 10$ and vary n.