02/10 (Th)

- grading: still on-going
- <u>PH</u> examples

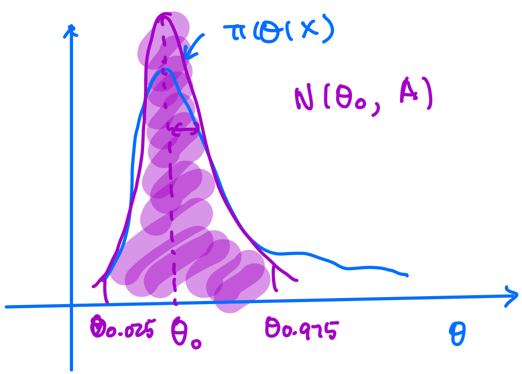
$\dagger$ Normal Approx. to Posterior - (1)

- General Idea: find a Gaussian approximation to $\pi(\theta \mid \boldsymbol{x})$.

- Consider a univariate case;

$$\underline{\pi(\theta \mid \boldsymbol{x})} = \frac{f(\boldsymbol{x} \mid \theta)\pi(\theta)}{m(\boldsymbol{x})} \propto \underline{q(\theta)}$$

$\star\star$ We find θ_0 and A such that $\pi(\theta \mid \boldsymbol{x}) \approx \mathrm{N}(\theta_0, A^{-1})$.



$E(\theta | x) \approx \theta_0$

θ_α s.t. $\Phi\left(\frac{\theta_\alpha - \theta_0}{\sqrt{A}}\right) = \alpha$

† Normal Approx. to Posterior - (2)

- θ_0: a mode of $\pi(\theta \mid \boldsymbol{x})$, i.e., a mode of $q(\theta)$.

$$\Rightarrow \text{ find } \theta_0 \text{ st } \frac{dq(\theta)}{d\theta} = 0.$$

⋆⋆ We can use any algorithms including numerical solution (e.g, Newton-Raphson method, R function `optim`).

† Normal Approx. to Posterior - (3)

- Compute a truncated Taylor expansion of $\log\{q(\theta)\}$ at mode θ_0,

$$\log\{q(\theta)\} \approx \log\{q(\theta_0)\} + \left.\frac{d\log\{q(\theta)\}}{d\theta}\right|_{\theta=\theta_0} \cancel{(\theta - \theta_0)}^{0} + \frac{1}{2}\left.\frac{d^2\log\{q(\theta)\}}{d\theta^2}\right|_{\theta=\theta_0}(\theta-\theta_0)^2.$$

⋆⋆ Let $A = -\left.\left(d^2\log\{q(\theta)\}/d\theta^2\right)\right|_{\theta=\theta_0}$ and we have

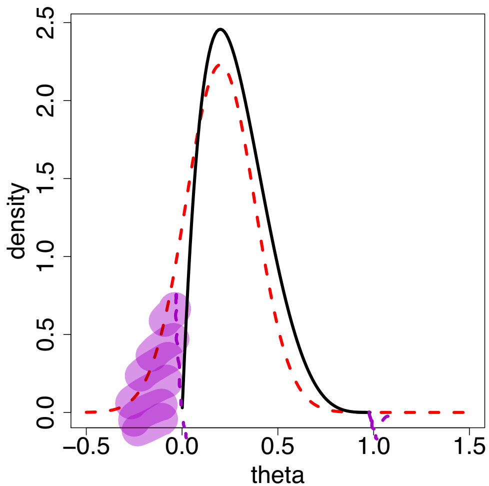
$$\log\{q(\theta)\} \approx \log\{q(\theta_0)\} - \frac{A}{2}(\theta-\theta_0)^2$$

$$\Rightarrow \quad q(\theta) \approx q(\theta_0)\exp\left\{-\frac{A}{2}(\theta-\theta_0)^2\right\}$$

$$\Rightarrow \quad \pi(\theta \mid \boldsymbol{x}) \approx \mathrm{N}(\theta_0, 1/A).$$

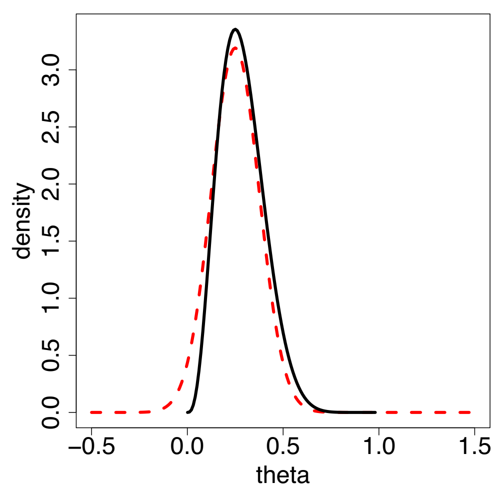
- **Example:** Suppose $\pi(\theta \mid \boldsymbol{x})$ is $\text{Be}(\alpha, \beta)$. The Laplace approximation gives us

$$\text{N}\left(\frac{\alpha-1}{\alpha+\beta-2}, \frac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^3}\right), \alpha, \beta > 1.$$



(a) $\alpha = 2$ & $\beta = 5$

$\alpha+\beta=7$



(b) $\alpha = 4$ & $\beta = 10$

$\alpha+\beta=14$

$\pi(\theta \mid x)$ is $Be(\alpha, \beta)$, $\theta \in (0, 1)$

$$\pi(\theta \mid x) = \frac{1}{B(\alpha, \beta)} \underbrace{\theta^{\alpha-1} (1-\theta)^{\beta-1}}$$

$$q(\theta) = \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

① Find θ_0

$$\log(q(\theta)) = (\alpha-1)\log\theta + (\beta-1)\log(1-\theta)$$

$$\frac{d \log(q(\theta))}{d\theta} = 0 \Rightarrow \text{find such } \theta$$

$$\theta_0 = \frac{\alpha-1}{\alpha+\beta-2}$$

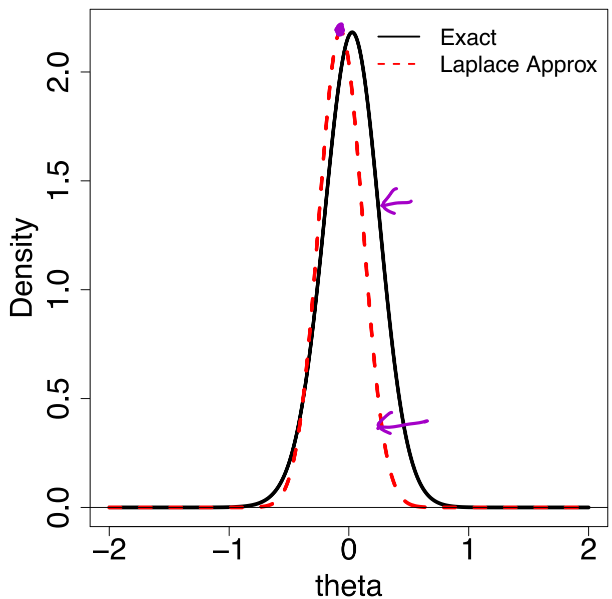
② Find A

$$A = -\frac{d^2 \log(q(\theta))}{d\theta^2}\Bigg|_{\theta=\theta_0} = \frac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^3}$$

approximate $Be(\alpha, \beta)$ with $N\left(\frac{\alpha-1}{\alpha+\beta-2}, \frac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^2}\right)$

- **Example:** Simulate a dataset of size $n = 15$, $x_i \overset{iid}{\sim} C(0, 1)$, $i = 1, \ldots n$.
Consider the estimation of the location of x and assume that $x_i \mid \theta \overset{iid}{\sim} C(\theta, 1)$ and $\theta \sim N(\mu, \sigma^2)$, with fixed $\mu = 0$ and $\sigma^2 = 25$.
We then approximate the posterior distribution of θ using the Laplace approximation.

$\pi(\theta | x) \propto \prod_{i=1}^{n} f(x_i | \theta) \; \pi(\theta)$



† Normal Approx. to Posterior - (4)

- Consider a multivariate case with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$;

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \approx \mathrm{N}(\boldsymbol{\theta}_0, A^{-1}),$$

where

⋆⋆ Find $\underline{\boldsymbol{\theta}_0} = (\theta_{0j}, j = 1, \ldots, p)$ such that $\frac{\partial q(\boldsymbol{\theta})}{\partial \theta_{0j}} = 0$.

⋆⋆ Find A, Hessian matrix evaluated at $\boldsymbol{\theta}_0$,

$$A_{ij} = -\frac{\partial^2 \log(q(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j}\Big|_{\theta=\theta_0}$$

- The Laplace approximation is only justified asymptotically – Smith et al (1985).

- The Laplace approximation seems to perform quite well in most cases (e.g.: the prior is smooth and the sample size is large) and can be useful as a guide to the solution of the problem.

- Normal approximations are not be useful if the posterior distributions are skewed or multimodal.

† Bayesian CLT

- Suppose $x_i \overset{iid}{\sim} f(x \mid \theta)$ where θ is a $p-$dim parameter and that the prior on θ is $\pi(\theta)$.

- Under some regularity conditions, the posterior probability distribution is approximately a normal distribution as sample size grows.

$$\pi(\theta \mid x) \rightarrow \mathrm{N}_p(\theta_0, A^{-1}), \text{ as } \ n \rightarrow \infty,$$

  where

  ⋆⋆ θ_0: posterior mode and A: Hassian matrix evaluated at θ_0.

- The prior can be improper, but assume that the posterior is proper.

⋆⋆ **Robert and Casella Example 3.16** (Gamma approximation)

As a simple illustration of the Laplace approximation for an integral, consider estimating a Gamma$(\alpha, 1/\beta)$ integral (mean: $\alpha\beta$),

$$\Pr(a < X < b) = \int_a^b f(x)dx = \int_a^b \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta} dx.$$

- check $x_0 = (\alpha - 1)\beta$ and $A = 1/(\alpha-1)/\beta^2$.
- Laplace approx. says

$\log(f(x)) \approx \log(f(x_0)) - \frac{A}{2}(x-x_0)^2$

$$f(x) \approx f(x_0)\sqrt{2\pi 1/A}\,\phi(x_0, 1/A),$$

pdf

$\frac{1}{\sqrt{2\pi 1/A}} e^{-\frac{A}{2}(x-x_0)^2}$

where $\phi(a, b)$ is the density function of $N(x_0, 1/A)$.

$$\Rightarrow \int_a^b f(x)dx \approx f(x_0)\sqrt{2\pi 1/A}\left\{\Phi(\sqrt{A}(b - x_0)) - \Phi(\sqrt{A}(a - x_0))\right\}.$$

⋆⋆ **Robert and Casella Example 3.16** (Gamma approximation – contd) Laplace approx. of a Gamma integral for $\alpha = 5$ and $\beta = 2$.

$(\alpha-1)\times\beta$

$= 4 \times 2 = 8$

Interval	Approximation	Exact
$(7, 9)$	0.193351	0.193341
$(6, 10)$	0.375046	0.37477
$(2, 14)$	0.848559	0.823349
$(15.987, \infty)$	0.0224544	0.100005

† Laplace Analytic Approximation (CR 6.2.3)

- Use the Laplace expansion to directly find

$$
\begin{aligned}
\mathrm{E}^{\pi}(g(\theta) \mid x) &= \frac{\int_{\Theta} g(\theta) f(x \mid \theta) \pi(\theta) d\theta}{\int_{\Theta} f(x \mid \theta) \pi(\theta) d\theta} \\
&= \frac{\int_{\Theta} \exp\{\tilde{q}^{\star}(\theta)\} d\theta}{\int_{\Theta} \exp\{\tilde{q}(\theta)\} d\theta},
\end{aligned}
$$

where $\tilde{q}^{\star}(\theta) = \log\{g(\theta) f(x \mid \theta) \pi(\theta)\}$ and $\tilde{q}(\theta) = \log\{f(x \mid \theta) \pi(\theta)\}$.

- Suppose $\tilde{q}^{\star}(\theta)$ and $\tilde{q}(\theta)$ have unique maxima, $\theta_0^{\star}$ and θ_0, respectively.

  ⋆⋆ Let $A^{\star} = -(d^2 \tilde{q}^{\star}(\theta)/d\theta^2)\Big|_{\theta=\theta_0^{\star}}$ and $A = -(d^2 \tilde{q}(\theta)/d\theta^2)\Big|_{\theta=\theta_0}$

- Then expand each in a second order Taylor expansion.

$$\mathrm{E}^{\pi}(g(\theta) \mid x) \quad = \quad \exp\{\tilde{q}^{\star}(\theta_0^{\star}) - \tilde{q}(\theta_0)\}\frac{\sqrt{A}}{\sqrt{A^{\star}}}.$$

- Can be extended for a multivariate $\boldsymbol{\theta}$.
- Lemma 6.2.4 and Corollary 6.2.5 discuss the Laplace approximation for $\mathrm{E}^{\pi}(g(\theta) \mid x)$. We skip them.

† Monte Carlo Method (PH 4)

- Suppose that we have $\theta^{(1)}, \ldots, \theta^{(M)}$ iid samples from $\pi(\theta \mid \boldsymbol{x})$.
- The law of large numbers implies that as $M \to \infty$,

⋆⋆ Posterior mean

$$\bar{\theta} = \frac{1}{M} \sum_{m=1}^{M} \theta^{(m)} \to \mathrm{E}(\theta \mid \boldsymbol{x}).$$

⋆⋆ Posterior variance

$$\frac{1}{M-1} \sum_{m=1}^{M} (\theta^{(m)} - \bar{\theta})^2 \to \mathrm{Var}(\theta \mid \boldsymbol{x}).$$

⋆⋆ Posterior probabilities

$$\frac{1}{M}\#(\theta^{(m)} \le c) \to \mathrm{P}(\theta \le c \mid \boldsymbol{x}).$$

⋆⋆ Posterior distribution function

$$\text{the empirical distribution of } \{\theta^{(1)}, \ldots, \theta^{(M)}\} \to \pi(\theta \mid \boldsymbol{x}).$$

⋆⋆ Posterior percentile

$$\text{the } \alpha\text{-percentile of } \{\theta^{(1)}, \ldots, \theta^{(M)}\} \to \theta_\alpha.$$

⋆⋆ Suppose $g(\theta) = \log(\theta/(1-\theta))$ for $0 < \theta < 1$

$$\frac{1}{M}\sum_{m=1}^{M} \log\left(\frac{\theta^{(m)}}{1-\theta^{(m)}}\right) \to \mathrm{E}\left(\log\left(\frac{\theta}{1-\theta}\right) \mid \boldsymbol{x}\right).$$

Similarly,

the empirical distribution of $\{g(\theta^{(1)}), \ldots, g(\theta^{(M)})\} \to \pi(g(\theta) \mid \boldsymbol{x})$.

⋆⋆ Posterior predictive distribution

$$p(x^{new} \mid \mathbf{x}) = \int p(x^{new} \mid \underline{\theta})\, \underbrace{\pi(\theta \mid \mathbf{x})}_{\theta^{(1)}, \ldots, \theta^{(M)},}\, d\theta$$

$$\text{sample } x_m^{new} \sim f(x \mid \theta^{(m)})$$

$x_1^{new}, \ldots, x_M^{new}$

The sequence of $\{x_m^{new}, \ldots, x_M^{new}\}$ constitutes M independent samples from the *marginal* posterior distribution of x.

⋆⋆ Go over Chapter 4 of PH for your practice.

† An illustration of Monte Carlo approximation: simulation study

⋆⋆ Suppose we have a dataset of size $n = 10$ with $x_i \in \mathbb{R}$, $i = 1, \ldots, n$. N(0,9)

⋆⋆ We consider the estimation problem of the mean of x. For the inference, we use a model that assumes

$$x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2) \text{ with fixed } \sigma^2 = 9,$$

and consider

$$\theta \sim \mathrm{N}(\mu, \tau^2) \text{ with } \mu = 0 \text{ and } \tau^2 = 2$$

for unknown θ.

† An illustration of Monte Carlo approximation: (contd)

⋆⋆ We can analytically obtain the posterior distributions of θ and of x^{new}

$$\theta \mid \boldsymbol{x} \sim \mathrm{N}(\mu_1, \tau_1^2), \quad \text{and} \quad x^{\text{new}} \mid \boldsymbol{x} \sim \mathrm{N}(\mu_1, \tau_1^2 + \sigma^2),$$

where $\tau_1^2 = (n/\sigma^2 + 1/\tau^2)^{-1}$ and $\mu_1 = \tau_1^2(n\bar{x}/\sigma^2 + \mu/\tau^2)$.

⇒ For our dataset, we obtained

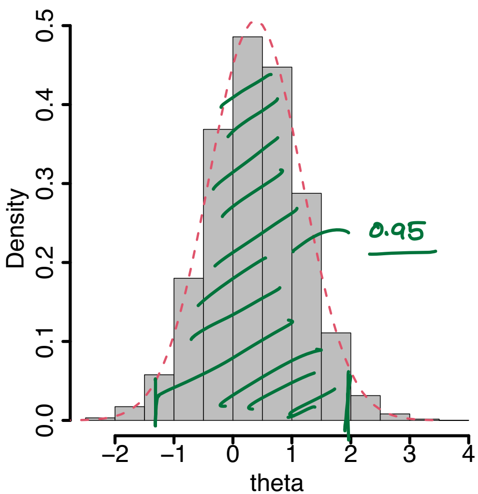
```
>
> c(post_m, post_var)
[1] 0.3834903 0.6206897
> c(pred_m, pred_var)
[1] 0.3834903 9.6206897
>
```

⋆⋆ Let's numerically approximate posterior quantities using the Monte Carlo method.

Simulate $\theta^{(m)}$ *independently* from $\mathrm{N}(\mu_1, \tau_1^2)$ and x_m^{new} from $\mathrm{N}(\theta^{(m)}, \sigma^2)$ $m = 1, \ldots, M$.

† An illustration of Monte Carlo approximation: (contd)

⋆⋆ The empirical distribution of $\{\theta^{(1)}, \ldots, \theta^{(M)}\}$ is a Monte Carlo approximation to $\pi(\theta \mid \boldsymbol{x})$.

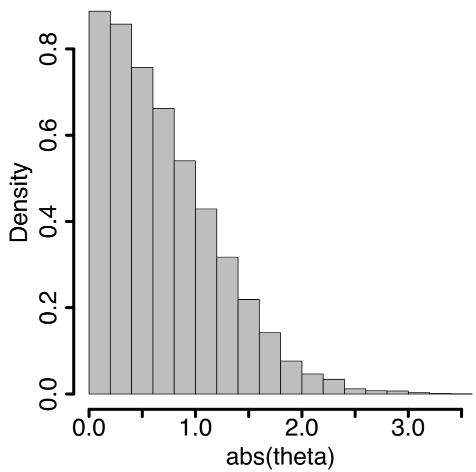


(a) $\pi(\theta \mid \boldsymbol{x})$

```
>
> mean(th_sam); post_m
[1] 0.3863699
[1] 0.3834903
> quantile(th_sam, prob=c(0.025, 0.5, 0.975))
      2.5%        50%      97.5%
-1.1696378  0.3880421  1.9103658
> qnorm(c(0.025, 0.5, 0.975), post_m, sqrt(post_var))
[1] -1.1606450  0.3834903  1.9276256
> var(th_sam); post_var
[1] 0.6253918
[1] 0.6206897
```

† An illustration of Monte Carlo approximation: (contd)

⋆⋆ Let $g(\theta) = |\theta|$.



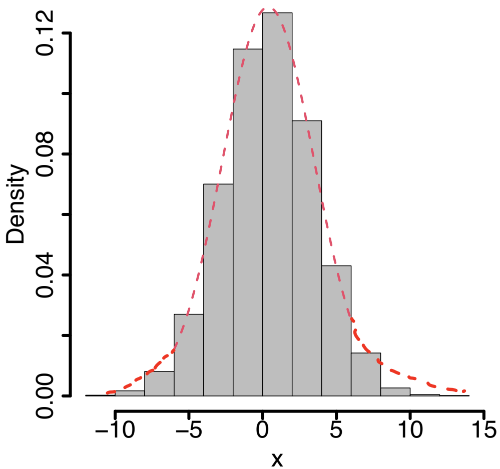
(a) $\pi(|\theta| \mid \boldsymbol{x})$

```
> mean(abs(th_sam));
[1] 0.7042984
> quantile(abs(th_sam), prob=c(0.025, 0.5, 0.975))
      2.5%        50%      97.5%
0.03009322 0.59956467 1.95162527
> var(abs(th_sam));
[1] 0.2786026
```

† An illustration of Monte Carlo approximation: (contd)

$\theta^{(m)} \rightarrow x_m^{new}$

$x_{m1}^{new} \; \cdots \; x_{mn}^{new}$



(a) $p(x^{\text{new}} \mid \boldsymbol{x})$

```
> mean(x_new); pred_m
[1] 0.4242155
[1] 0.3834903
> quantile(x_new, prob=c(0.025, 0.5, 0.975))
      2.5%        50%      97.5%
-5.6653821  0.4257596  6.4333582
> qnorm(c(0.025, 0.5, 0.975), pred_m, sqrt(pred_var))
[1] -5.6957764  0.3834903  6.4627569
> var(x_new); pred_var
[1] 9.553585
[1] 9.62069
```

★★ PH Chapter 4 has a thorough example of a gamma distribution.

† Simulating Samples from Distributions

- Most statistical packages provide random number generators to simulate from common families of distributions, e.g.,

```
> runif(1, 0, 1)
[1] 0.985409
```

† Simulating Samples from Distributions (contd)

- Starting with samples from the uniform distribution Unif(0,1), we can generate samples from various distributions through transformations. e.g.,

  ⋆⋆ If $U \sim \text{Unif}(0,1)$, then $W = -\log(U) \sim \text{Exp}(1)$ and $V = \lambda W \sim \text{Gamma}(1, \lambda)$.

  ⋆⋆ If $U_1, U_2 \overset{iid}{\sim} \text{Unif}(0,1)$, we obtain a pair of indep. standard normal random variables $(Z_1, Z_2) = (\sqrt{-2\log(U_1)}\cos(2\pi U_2), \sqrt{-2\log(U_1)}\sin(2\pi U_2))$ by the Box-Muller transformation.

† Simulating Samples from Distributions (contd)

- Inverse CDF Method: Use the probability integral transform $U = F(X) = \int_{-\infty}^{x} f(s)ds$.

  ⋆⋆ We can easily see U have a Unif$(0, 1)$ distribution.

  ⋆⋆ So generate X having cdf F via $X = F^{-1}(U)$ (works nicely when F^{-1} has a simple analytic form).

- e.g., let X have Exp(λ), i.e., $F(x) = 1 - e^{-\lambda x}$.

  ⋆⋆ Generate $U \sim$ Unif$(0, 1)$ and let $X = -\log(1 - U)/\lambda$.

- Also, check rejection sampling, adaptive rejection sampling...

† Monte Carlo Integration – Importance Sampling

- Recall that we have a problem of approximating

$$\pi(\theta \mid x)$$

$$E^{\pi}(g(\theta) \mid \boldsymbol{x}) = \frac{\int_\Theta g(\theta) f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta) d\theta}.$$

- We can actually generate $(\theta^{(1)}, \ldots, \theta^{(M)})$ from a density other than the distribution function of interest and approximate the integral.

- Suppose h is a probability density function with supp(h) including the support of $g(\theta)f(x \mid \theta)\pi(\theta)$.

- We have

$$\mathrm{E}(g(\theta) \mid \boldsymbol{x}) = \frac{\int_\Theta g(\theta) f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta}{\int_\Theta f(\boldsymbol{x} \mid \theta)\pi(\theta)d\theta}$$

$\theta^{(m)} \overset{iid}{\sim} h(\theta)$

- We express

$$\begin{aligned}
\int_\Theta g(\theta) f(x \mid \theta)\pi(\theta)d\theta &= \int_\Theta \frac{g(\theta) f(x \mid \theta)\pi(\theta)}{h(\theta)} h(\theta)d\theta \\
\int_\Theta f(x \mid \theta)\pi(\theta)d\theta &= \int_\Theta \frac{f(x \mid \theta)\pi(\theta)}{h(\theta)} h(\theta)d\theta.
\end{aligned}$$

- The method of *importance sampling* is an evaluation of the integral based on generating a sample $\theta^{(1)}, \ldots, \theta^{(M)}$ from a given distribution $h(\theta)$ and approximating

$$\begin{aligned}
\int_\Theta g(\theta) f(\mathbf{x} \mid \theta)\pi(\theta) d\theta &\approx \frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) \frac{f(\mathbf{x} \mid \theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})} \\
&= \frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) w_m, \\
\int_\Theta f(\mathbf{x} \mid \theta)\pi(\theta) d\theta &\approx \frac{1}{M}\sum_{m=1}^{M} w_m.
\end{aligned}$$

⋆⋆ $h(\theta)$: importance function

⋆⋆ $w_m = w(\theta^{(m)}) = \frac{f(\mathbf{x}|\theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})}$: weights

f

$$\Rightarrow \quad \mathrm{E}(g(\theta) \mid \mathbf{x}) = \frac{\int_\Theta g(\theta) f(\mathbf{x} \mid \theta)\pi(\theta) d\theta}{\int_\Theta f(\mathbf{x} \mid \theta)\pi(\theta) d\theta} \approx \frac{\frac{1}{M}\sum_{m=1}^{M} g(\theta^{(m)}) w_m}{\frac{1}{M}\sum_{m=1}^{M} w_m}.$$

- **Example** (Example 6.1.1 with some changes)

  ⋆⋆ Consider a dataset $\boldsymbol{x} = (x_1, \ldots, x_n)$ with $n = 10$, where x_i's are simulated from $\mathrm{N}(0, 9)$.

  ⋆⋆ We consider the estimation problem of the mean of x. For the inference, we use a model that assumes

  $$x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2) \text{ with fixed } \sigma^2 = 9,$$

  and consider

  $$\theta \sim \mathrm{N}(\mu, \tau^2) \text{ with } \mu = 0 \text{ and } \tau^2 = 2$$

  for unknown θ.

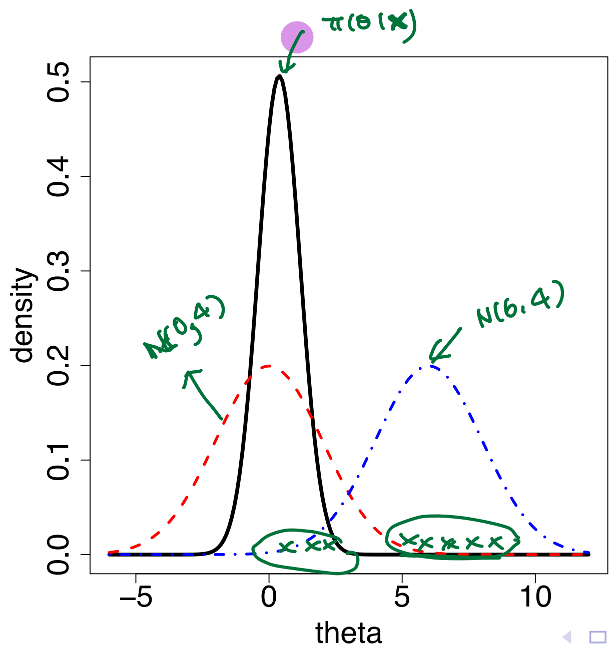
  ⋆⋆ Suppose we use the Bayes estimator of $g(\theta) = \theta$ under the squared error loss

  $$\delta^{\pi}(\boldsymbol{x}) = \mathrm{E}(\theta \mid \boldsymbol{x}) = \left(\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}\right)^{-1} \left(\frac{\bar{x}}{\sigma^2/n} + \frac{\mu}{\tau^2}\right).$$

- **Example** (contd)
  - For my simulated dataset, the exact value of $\delta^\pi(\boldsymbol{x}) = 0.38349$
  - Let's use the importance sampling method to numerically approximate $\delta^\pi(\boldsymbol{x})$;

⋆⋆ Case 1: $h(\theta) = N(0, 2^2)$

⋆⋆ Case 2: $h(\theta) = N(6, 2^2)$



- Generate $\theta^{(m)}$, $m = 1, \ldots, M$ from $\overbrace{\mathrm{N}(a, v^2)}^{h(\theta)}$ for large enough M.

- Compute

$$\rightarrow \hat{\delta}^{\pi}(\boldsymbol{x}) = \frac{\sum_{m=1}^{M} \theta^{(m)} w_m}{\sum_{m=1}^{M} w_m},$$

where $w_m = w(\theta^{(m)}) = \dfrac{f(\boldsymbol{x}|\theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})}$.

```
>
> c(post_mean_0, post_mean_1, post_mean_2)
[1] 0.3834903 0.3885833 0.2876347
```

$\delta^{\pi}(x)$

IS approx w/ choice 1

IS approx w/ choice 2

- **Example** (contd)

$\delta^{\pi}(x)$

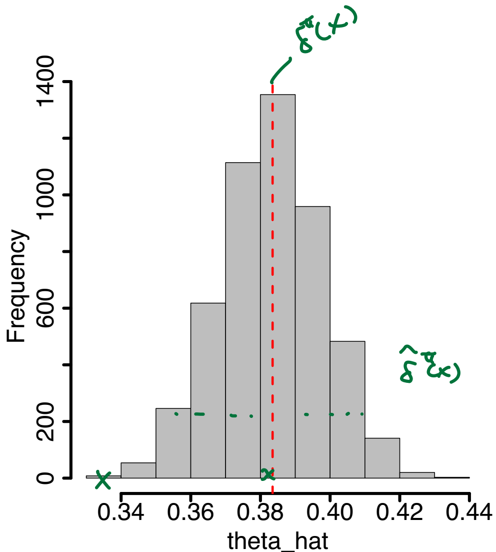
$\pi(\theta|x)$

$*$ Repeat 5000 times and make histograms of approximated $\delta^{\pi}(\boldsymbol{x})$ for the same dataset.

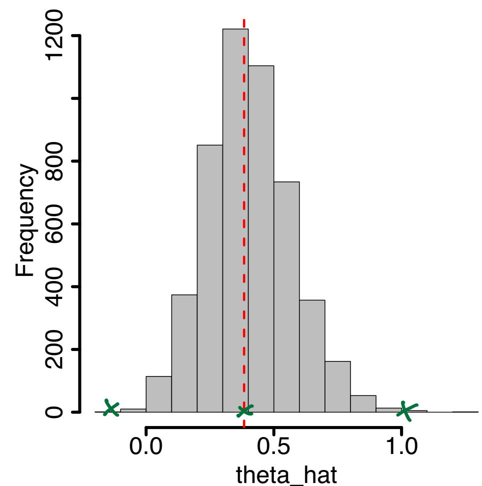
$\theta^{(1)}, \ldots, \theta^{(M)} \sim h(\theta) \Rightarrow \hat{\delta}^{\pi}(x)$

$\theta^{(1)}, \ldots, \theta^{(M)} \sim h(\theta) \Rightarrow \hat{\delta}^{\pi}(x)$

$\vdots$



(a) $h(\theta) = \mathrm{N}(0, 2^2)$



(b) $h(\theta) = \mathrm{N}(6, 2^2)$

- **Example** - Case 1: $h(\theta) = \mathrm{N}(0, 2^2)$

```
> summary(imp_v)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.000e+00 7.821e-14 9.000e-13 1.343e-12 2.580e-12 3.594e-12
```

- Case 2: $h(\theta) = \mathrm{N}(6, 2^2)$

```
> summary(imp_v)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
0.000e+00 0.000e+00 0.000e+00 1.592e-12 0.000e+00 3.743e-10
```

⋆⋆ Recall: $w_m = w(\theta^{(m)}) = \frac{f(x|\theta^{(m)})\pi(\theta^{(m)})}{h(\theta^{(m)})}$: weights.

- Remarks

  ⋆⋆ Simulation according to h must be easily implemented, requiring a fast and reliable pseudo-random generator.

  ⋆⋆ h can be almost any density but the choice of the importance function h is crucial.

  ⋆⋆ The function $h(\theta)$ must be close enough to $g(\theta)\pi(\theta)$ to reduce the variability of $\hat{\mathrm{E}}^{\pi}(g(\theta) \mid x)$.

  ⋆⋆ Obviously there are some choices that are better than others, and it is natural to try to compare different distinctions h for the evaluation of $\mathrm{E}^{\pi}(g(\theta) \mid x)$.

$\theta^{(0)}$ $\theta^{(t)} \sim K(\cdot \mid \theta^{(t-1)})$, $t=1, \ldots\ldots$ $\theta^{(t)} \sim \pi(\theta \mid x)$ Θ

$\theta^{(1)}, \ldots, \theta^{(T)}$

† Markov chain Monte Carlo (MCMC) methods (CR 6.3)

- *A more general Monte Carlo method* that approximates the generation of random variables from $\pi(\theta \mid x)$.

irreducible
positive recurrent
aperiodic

Ergodic Markov chain

- A Markov chain is a sequence of random variables $\theta^{(1)}, \theta^{(2)}, \ldots$, where for any t, the distribution of $\theta^{(t)}$ given all previous θ's depends only on the most recent value, $\theta^{(t-1)}$.

  i.e., draw $\theta^{(t)}$ from a transition distribution (the transition kernel of the Markov chain), $K(\theta^{(t)} \mid \theta^{(t-1)})$.

- If $K(\cdot \mid \cdot)$ satisfies certain conditions (*detailed balance condition*), the distribution of $\theta^{(t)}$ converges to a unique stationary distribution that is the posterior distribution as t grows, regardless of where the chain was initiated.