posterior mode: Suppose the posterior distribution of θ is $f(\theta \mid x_1, \ldots, x_n)$. The posterior mode of θ is defined as the point where $f(\theta \mid x_1 \ldots x_n)$ takes the maximum, i.e., the point $\hat{\theta}$ such that

$$f(\hat{\theta} \mid x_1 \ldots x_n) \geq f(\theta \mid x_1 \ldots x_n), \text{ for any } \theta.$$

Way to find the posterior mode:

step 1. let $l(\theta) = \log f(\theta \mid x_1 \ldots x_n)$.

step 2. $\frac{\partial}{\partial \theta} l(\theta) = 0$. get $\hat{\theta}$ as the solution.

Eg 5. $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma_0^2)$. σ_0^2 is known. let the prior on μ be $\mu \sim N(\mu_0, \tau_0^2)$. where μ_0 and τ_0^2 are known.

then the posterior distribution of μ is

$$\mu \mid X_1 \ldots X_n \sim N(\hat{\mu}, \sigma^2). \text{ where}$$

$$\hat{\mu} = \frac{n\bar{x}\tau_0^2 + \mu_0\sigma_0^2}{n\tau_0^2 + \sigma_0^2}, \quad \sigma^2 = \frac{\sigma_0^2 \tau_0^2}{n\tau_0^2 + \sigma_0^2}.$$

So $f(\mu \mid x_1 \ldots x_n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(\mu-\hat{\mu})^2}{2\sigma^2}}$.

$l(\mu) = \log f(\mu \mid x_1 \ldots x_n) = \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{(\mu-\hat{\mu})^2}{2\sigma^2}$. So $\frac{\partial}{\partial \mu} l(\mu) = -\frac{(\mu - \hat{\mu})}{\sigma^2} = 0$

So, $\hat{\mu} = \dfrac{n\bar{x}\tau_0^2 + \mu_0\sigma_0^2}{n\tau_0^2 + \sigma_0^2}$ is the posterior mode for μ. the same as the posterior mean!

Eg 6. If $X_1, \dots, X_n$ are $\overset{iid}{\sim}$ Bernoulli (θ). where $\theta \sim$ Beta(α, β).

both $\alpha, \beta > 0$ are known. It follows from. Eg 2. that

the posterior distribution of θ is

$$\theta \mid X_1, \dots, X_n \sim \text{Beta}(n\bar{x} + \alpha,\ n(1-\bar{x}) + \beta).$$

so $$f(\theta \mid x_1, \dots, x_n) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(n\bar{x}+\alpha)\,\Gamma(n(1-\bar{x})+\beta)}\ \theta^{n\bar{x}+\alpha-1}(1-\theta)^{n(1-\bar{x})+\beta-1}.$$

$$l(\theta) = \log \frac{\Gamma(n+\alpha+\beta)}{\Gamma(n\bar{x}+\alpha)\,\Gamma(n(1-\bar{x})+\beta)} + (n\bar{x}+\alpha-1)\log\theta + (n(1-\bar{x})+\beta-1)\log(1-\theta).$$

so $$\frac{\partial}{\partial\theta} l(\theta) = \frac{n\bar{x}+\alpha-1}{\theta} - \frac{n(1-\bar{x})+\beta-1}{1-\theta} = 0$$

we get the posterior mode is

$$\hat{\theta} = \frac{n\bar{x}+\alpha-1}{n+\alpha+\beta-2}.$$

Different from the posterior mean!

Eg7.* If $(x_1, \cdots, x_k)$ follows multinomial model. i.e.

$$(x_1, \cdots, x_k) \sim \text{multinomial}(n; \theta_1, \cdots, \theta_k).$$

for $\underbrace{\theta_1, \cdots, \theta_k}^{\text{unknown}} \geq 0$, $\theta_1 + \cdots + \theta_k = 1$.

consider Dirichlet prior on θ ~~then the~~. that is

$$\theta \sim \text{Dirichlet}(\alpha_1, \cdots, \alpha_k). \quad \alpha_1 \cdots \alpha_k > 0 \text{ are known}$$

Then, the posterior distribution of θ is

$$\theta | x_1 \cdots x_k \sim \text{Dirichlet}(x_1 + \alpha_1, \cdots, x_k + \alpha_k).$$

So $f(\theta_1, \cdots, \theta_k | x_1 \cdots x_k) \propto \theta_1^{x_1+\alpha_1-1} \theta_2^{x_2+\alpha_2-1} \cdots \theta_k^{x_k+\alpha_k-1}$.

So, $l(\theta_1, \cdots \theta_k) = \log f(\theta_1, \cdots \theta_k | x_1, \cdots x_k)$

$$= (x_1+\alpha_1-1)\log\theta_1 + (x_2+\alpha_2-1)\log\theta_2 + \cdots + (x_k+\alpha_k-1)\log\theta_k.$$

$$= (x_1+\alpha_1-1)\log\theta_1 + \cdots + (x_{k-1}+\alpha_{k-1}-1)\log\theta_{k-1}$$
$$+ (x_k+\alpha_k-1)\log(1-\theta_1-\cdots-\theta_{k-1})$$

So $\frac{\partial}{\partial\theta} l(\theta) = 0$. we have

$$\begin{cases} \dfrac{x_1+\alpha_1-1}{\theta_1} - \dfrac{x_k+\alpha_k-1}{1-\theta_1-\theta_2-\cdots-\theta_{k-1}} = 0. \\ \vdots \\ \dfrac{x_{k-1}+\alpha_{k-1}-1}{\theta_{k-1}} - \dfrac{x_k+\alpha_k-1}{1-\theta_1-\theta_2-\cdots-\theta_{k-1}} = 0. \end{cases}$$

In other words.

$$\frac{x_1+\alpha_1-1}{\theta_1} = \frac{x_2+\alpha_2-1}{\theta_2} = \cdots = \frac{x_{k-1}+\alpha_{k-1}-1}{\theta_{k-1}} = \frac{x_k+\alpha_k-1}{1-\theta_1-\theta_2-\cdots-\theta_{k-1}}$$

$$= \frac{x_1+\cdots+x_k+\alpha_1+\alpha_2+\cdots+\alpha_k-k.}{\theta_1+\theta_2+\cdots+\theta_{k-1}+1-\theta_1-\cdots-\theta_{k-1}} = x_1+\cdots+x_k+\alpha_1+\cdots+\alpha_k-k.$$

$$= n+\alpha_1+\cdots+\alpha_k-k$$

So $\begin{cases} \hat{\theta}_1 = \dfrac{x_1+\alpha_1-1}{n-k+\alpha_1+\cdots+\alpha_k.} \\ \vdots \\ \hat{\theta}_k = \dfrac{x_k+\alpha_1-1}{n-k+\alpha_1+\cdots+\alpha_k} \end{cases}$ is the posterior mode.

EX4. Recall that the MLE of θ is

$$\begin{cases} \theta_1 = \dfrac{x_1}{n}, \\ \vdots \\ \theta_k = \dfrac{x_k}{n}. \end{cases}$$

Find values of $\alpha_1, \cdots, \alpha_k$ s.t. the above posterior mode is the same as the MLE.

Sol: let $\begin{cases} \dfrac{x_1+\alpha_1-1}{n-k+\alpha_1+\cdots+\alpha_k} = \dfrac{x_1}{n.} \\ \vdots \\ \dfrac{x_k+\alpha_k-1}{n-k+\alpha_1+\cdots+\alpha_k} = \dfrac{x_k}{n.} \end{cases}$ So. $\alpha_1 = \cdots = \alpha_k = 1$

§7.2.2. Credible interval.

Eg 1. (Normal model.). Let $X_1, X_2, \cdots X_n$ be iid samples from $N(\mu, \sigma_0^2)$, where $\sigma_0^2 > 0$ is known. Assume normal prior on μ, i.e., $\mu \sim N(\mu_0, \tau_0^2)$, where μ_0, τ_0^2 are known.

Then the posterior distribution of μ is

$$\mu | X_1, \cdots, X_n \sim N(\hat{\mu}, \sigma^2).$$

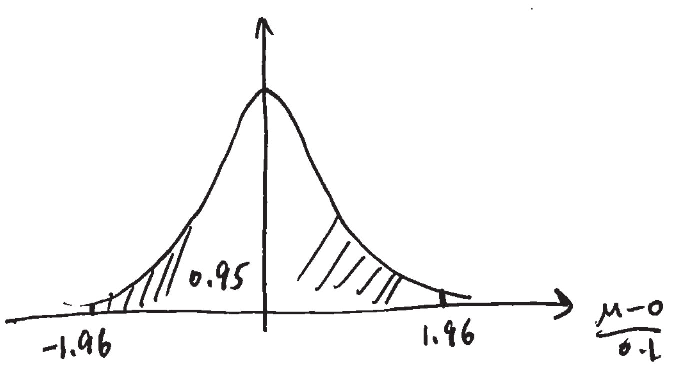
where $\hat{\mu} = \dfrac{n\bar{X}\tau_0^2 + \sigma_0^2\mu_0}{n\tau_0^2 + \sigma_0^2}$, $\sigma^2 = \dfrac{\sigma_0^2 \tau_0^2}{n\tau_0^2 + \sigma_0^2}$.

If $n = 99$, $\bar{X} = \mu_0 = 0$, $\sigma_0^2 = \tau_0^2 = 1$, then

$$\hat{\mu} = 0, \quad \sigma^2 = 0.01.$$

So $\mu | X_1 \cdots X_n \sim N(0, 0.01)$.

So. $\dfrac{\mu - 0}{0.1} \sim Z = N(0,1)$, given $X_1 \cdots X_n$.



So $P\left(-1.96 < \dfrac{\mu - 0}{0.1} < 1.96 \,\middle|\, X_1 \cdots X_n\right) = 0.95$

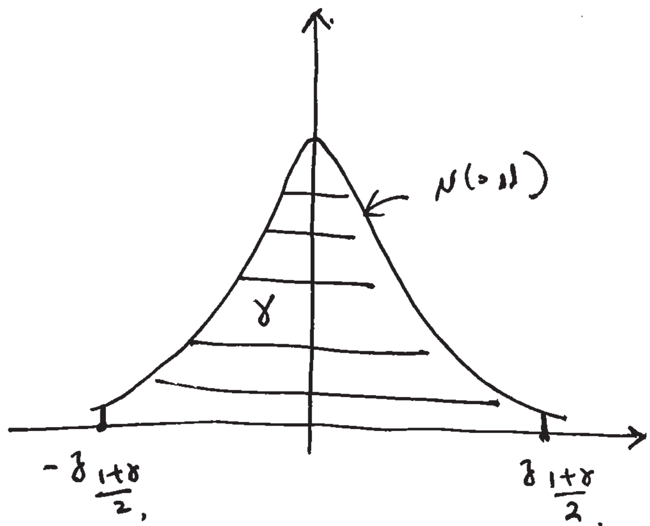
So the 95% - credible interval is

$$-0.196 < \mu < 0.196$$

In general, given the samples $X_1 \text{---} X_n$.

$$\frac{\mu - \hat{\mu}}{\sigma} \sim N(0,1).$$

so $\frac{\mu - \hat{\mu}}{\sigma}$ follows the standard normal curve



So $$p\left(-z_{\frac{1+\gamma}{2}} < \frac{\mu - \hat{\mu}}{\sigma} < z_{\frac{1+\gamma}{2}} \;\middle|\; X_1 \cdots X_n \right) = \gamma.$$

The γ - credible interval is

$$\hat{\mu} - z_{\frac{1+\gamma}{2}} \cdot \sigma, \quad < \mu < \quad \hat{\mu} + z_{\frac{1+\gamma}{2}} \cdot \sigma$$

or $$\hat{\mu} \pm z_{\frac{1+\gamma}{2}} \cdot \sigma$$

EX. ~~4~~ In Eg1. if $n = 15$, $\mu_0 = 0$. $\tau_0^2 = \sigma_0^2 = 1$. $\bar{x} = 1$.

① find 90% credible level for μ.

sol.
$$\hat{\mu} = \frac{n\bar{x}\tau_0^2 + \sigma_0^2\mu_0}{n\tau_0^2 + \sigma_0^2} = \frac{15}{16}.$$

$$\sigma^2 = \frac{\sigma_0^2\tau_0^2}{n\tau_0^2 + \sigma_0^2} = \frac{1}{16}.$$

So. the 90% credible interval for μ is.

$$\hat{\mu} \pm 1.645\,\sigma = \frac{15}{16} \pm 1.645 \cdot \frac{1}{4}. = (0.52625,\ 1.34875).$$

~~Ex 4, In EX3~~, ②. How many samples are needed such that the credible interval is shorter than 0.2.

Sol. The length of the credible interval is

$$2 \times z_{\frac{1+\gamma}{2}}\ \sigma,$$

For $\gamma = 0.9$, this becomes $2 \times 1.645 \times \frac{1}{\sqrt{n+1}}$.

let
$$2 \times 1.645 \times \frac{1}{\sqrt{n+1}} \leq 0.2$$

then. $n \geq 269.6025$. or $n \geq 270$.

In general, ~~set~~ to make the credible ~~[illegible]~~ interval shorter than δ.

just set

$$2 z_{\frac{\gamma+1}{2}} \sqrt{\frac{\sigma_0^2 \tau_0^2}{n\tau_0^2 + \sigma_0^2}} \leq \delta .$$

we get $$n \geq \frac{4 z_{\frac{\gamma+1}{2}}^2 \tau_0^2 \sigma_0^2 - \sigma_0^2 \delta^2}{\tau_0^2 \delta^2} .$$