

1. (a) **Given information:**  $\theta|y \sim N(\mu, \sigma^2)$ , have  $n$  independent samples of  $\theta$ .  $\text{sd}(\theta|y) = \sigma$

To estimate the  $p\%$  quantile from a sample  $\theta^1, \theta^2, \dots, \theta^n$ , you sort the sample to obtain  $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$ .  $\theta^{(k)}$  is called the *kth order statistic* and refers to the  $k$ th smallest item in the sample. The estimate for the  $p\%$  quantile is then  $\theta^{(\lfloor np \rfloor)}$  (you may interpolate between the two nearest integers around  $np$ , but here we are concerned about the accuracy so we won't discuss this).

The variance of this estimator is:

$$\sigma_{est}^2 = \frac{p(1-p)}{nf(F^{-1}(p))^2} \quad (1)$$

Where  $f$  is the PDF and  $F^{-1}$  is the inverse CDF. That is,  $F^{-1}(p)$  gives you the  $p\%$  quantile. The  $p(1-p)/n$  term comes from the variance of binomial distribution (sum of bernoulli indicators) as demonstrated in lecture 4.1. I don't know the justification behind the  $f$  term and would appreciate for someone to shed light on that.

This equation seems unhelpful, since we would like to estimate the variance, but we need the density of the distribution we're trying to estimate! Luckily, we only want to control the deviation proportional to the standard deviation of the posterior, which we can compute for the normal distribution in particular.

If you work out the math, you'll find that:

$$f_{N(\mu, \sigma^2)}(F_{N(\mu, \sigma^2)}^{-1}(p)) = \frac{f_{N(0,1)}(F_{N(0,1)}^{-1}(p))}{\sigma} \quad (2)$$

Which is now only depends on our  $\sigma$ . Plugging this into (1):

We are concerned about  $p = 2.5\%$  and  $p = 97.5\%$ :

$$f_{N(0,1)}(F_{N(0,1)}^{-1}(0.025))^2 = f_{N(0,1)}(-1.96)^2 = 0.003416$$

By symmetry, we get the same value for  $p = 97.5\%$ . We now solve for the minimum  $n$ :

$$0.1\sigma \geq \sqrt{\frac{p(1-p)\sigma^2}{0.003416n}} \implies n \geq \frac{p(1-p)}{0.00003416}$$

Plugging in  $p = 2.5\%$  and  $p = 97.5\%$ , we know we require  $n \geq 714$  for an accuracy of  $0.1\text{sd}(\theta|y)$ .

2. **Given information:** Have 100 independent draws of  $\theta_1$ , which is approximately  $N(8, 4^2)$ .

(a) By central limit theorem, the standard deviation of the sample mean is:

$$\sigma_{\text{sim}} = \sqrt{\frac{\text{Var}(\theta_1)}{n}} = \sqrt{\frac{4^2}{100}} = 0.4$$

Thus the standard deviation due to simulation variability is 0.4. Notice we do not incorporate the epistemic uncertainty (i.e.  $\sigma$ ) here because we only want the standard deviation *due to simulation variability*.

(b) To reduce  $\sigma_{\text{sim}}$  to 0.1, we want need a reduction by a factor of 4:

$$\frac{\sigma_{\text{sim}}}{4} = \frac{1}{4} \sqrt{\frac{\text{Var}(\theta_1)}{n}} = \sqrt{\frac{\text{Var}(\theta_1)}{16n}}$$

Our original  $n = 100$ , thus we need 1600 simulation draws for a simulation standard deviation of 0.1.

- (c) In Q1, we computed the *total* standard deviation of the estimate. Here, we are concerned with the *simulation* standard deviation of the estimate. What is the simulation standard deviation for a quantile?
- (d) Depends on (c)
- (e) Skipped due to not covering this example yet.
- (f) Skipped due to not covering this example yet.