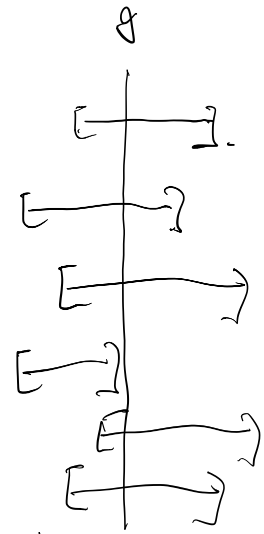
Confidence intervals.

$$P_\theta \longrightarrow S_n^{(0)} = \{X_1^{(0)}, \ldots, X_n^{(0)}\} \longrightarrow C_n^{(0)} = C_n(S_n^{(0)})$$

$$S_n^{(1)} = \{X_1^{(1)}, \ldots, X_n^{(1)}\} \rightarrow C_n^{(1)}$$

$$S_n^{(2)} = \{X_1^{(2)}, \ldots, X_n^{(2)}\} \rightarrow C_n^{(2)}$$

$$\vdots$$



We say that C_n has 95% coverage if with respect to 95% of hypothetical datasets generated from P_θ, it contains the true parameter θ.

Rem: θ is fixed, C_n is random.

Credible intervals.

$$X = \{X_1, \ldots, X_n\}.$$
$$+$$
$$\text{prior} \implies \text{posterior } p(\theta | X).$$

Def: A credible interval with 95% coverage is an interval $C_n = [a, b]$ such that

$$\int_a^b p(\theta|x)\,d\theta = P_{p(\theta|x)}(C_n) \geq 0.95.$$

Rem: The interval C_n is fixed, θ is random.

For Bayesian supervised learning, we immediately get uncertainty quantification for labels.

Using the training data, have posterior

$$p(\beta \mid X_{train}, y_{train}).$$

Observe a new sample X_{new}, make a prediction

$$\hat{y}_{new} = X_{new}^T \beta_{MAP}.$$

Also, have the entire posterior predictive distribution

$$p(y_{new} \mid X_{train}, y_{train}, X_{new}).$$

Use this to get 95% credible interval for y_{new}.

(i.e. interval C_{new} such that

$$\mathbb{P}_{p(y_{new} \mid X_{train}, y_{train}, X_{new})}(y_{new} \in C_{new}) \geq 0.95 \quad).$$

In frequentist supervised learning, also want to get a 95% confidence interval for the label y of a sample X.

Note: There are 2 sources of randomness.

① Error in estimate $\mathbb{E}[Y \mid X = x]$.

② Randomness in the sample label, i.e. $Y \mid X = x$ is random.

These are called prediction intervals.

Performance metrics for confidence / credible / prediction intervals.

1. High coverage.
2. Low width.

There is a tradeoff between high coverage and low width, i.e. the higher the coverage, the larger the width of the ___ intervals you need.

Validity

We say that a CI with coverage $1-\alpha$ is valid if it has the coverage that is advertized.

There are 2 types of validity.

1. Exact validity: Advertized coverage holds in finite samples.
2. Asymptotic validity: Advertized coverage holds asymptotically (i.e. as no. of samples $n \to \infty$).

Example 1:

Suppose we draw $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$. with σ^2 known.

Goal: Estimate mean μ.

Then a 95% CI for μ is given by

$$\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right).$$

This is exactly valid.

Pf: $\mathbb{P}\left(\mu \in \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)\right)$.

$= \mathbb{P}\left(|\mu - \bar{X}| \le 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right)$. $\qquad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$= \mathbb{P}\left(\frac{\sqrt{n}|\bar{X} - \mu|}{\sigma} \le 1.96\right)$. $\qquad \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$.

≥ 0.95. ∎

Example 2: Consider $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathrm{Unif}([\mu - \frac{1}{2}, \mu + \frac{1}{2}])$

Goal: Estimate μ.

Know $\mathrm{Var}(\bar{X}) = \frac{1}{4n}$, so $2\sqrt{n}(\bar{X} - \mu) \Rightarrow N(0, 1)$.

Hence, $C_n = \left(\bar{X} - \frac{1.96}{2\sqrt{n}},\ \bar{X} + \frac{1.96}{2\sqrt{n}}\right)$.

Then $\lim_{n\to\infty} \mathbb{P}(\mu \in C_n) = \lim_{n\to\infty} \mathbb{P}\left(|\bar{X} - \mu| \le \frac{1.96}{2\sqrt{n}}\right)$.

$= \lim_{n\to\infty} \mathbb{P}(\underbrace{2\sqrt{n}|\bar{X} - \mu|} \le 1.96)$.

≥ 0.95.

Hence, C_n is an asymptotically valid 95% CI, but not an exactly valid 95% CI.