2. Pivot $R_n = \widehat{\theta}_n - \theta$
3. Let $H(r) = \mathbb{P}[R_n \le r]$ be the CDF of R_n
4. Let $R^*_{n,b} = \widehat{\theta}^*_{n,b} - \widehat{\theta}_n$. Approximate H using bootstrap:

$$\widehat{H}(r) = \frac{1}{B}\sum_{b=1}^{B} I(R^*_{n,b} \le r)$$

5. $\theta^*_\beta = \beta$ sample quantile of $(\widehat{\theta}^*_{n,1}, \dots, \widehat{\theta}^*_{n,B})$
6. $r^*_\beta =$ beta sample quantile of $(R^*_{n,1}, \dots, R^*_{n,B})$, i.e., $r^*_\beta = \theta^*_\beta - \widehat{\theta}_n$
7. Approximate $1-\alpha$ confidence interval $C_n = \left(\hat{a}, \hat{b}\right)$ where

$$\begin{aligned} \hat{a} &= \quad \widehat{\theta}_n - \widehat{H}^{-1}\left(1 - \frac{\alpha}{2}\right) = \quad \widehat{\theta}_n - r^*_{1-\alpha/2} = \quad 2\widehat{\theta}_n - \theta^*_{1-\alpha/2} \\ \hat{b} &= \quad \widehat{\theta}_n - \widehat{H}^{-1}\left(\frac{\alpha}{2}\right) = \quad \widehat{\theta}_n - r^*_{\alpha/2} = \quad 2\widehat{\theta}_n - \theta^*_{\alpha/2} \end{aligned}$$

Percentile interval

$$C_n = \left(\theta^*_{\alpha/2}, \theta^*_{1-\alpha/2}\right)$$

16.3 Rejection Sampling

Setup

- We can easily sample from $g(\theta)$
- We want to sample from $h(\theta)$, but it is difficult
- We know $h(\theta)$ up to a proportional constant: $h(\theta) = \dfrac{k(\theta)}{\int k(\theta)\, d\theta}$
- Envelope condition: we can find $M > 0$ such that $k(\theta) \le M g(\theta) \quad \forall\theta$

Algorithm

1. Draw $\theta^{cand} \sim g(\theta)$
2. Generate $u \sim \text{Unif}\,(0,1)$
3. Accept θ^{cand} if $u \le \dfrac{k(\theta^{cand})}{Mg(\theta^{cand})}$
4. Repeat until B values of θ^{cand} have been accepted

Example

- We can easily sample from the prior $g(\theta) = f(\theta)$
- Target is the posterior $h(\theta) \propto k(\theta) = f(x^n \mid \theta) f(\theta)$
- Envelope condition: $f(x^n \mid \theta) \le f(x^n \mid \widehat{\theta}_n) = \mathcal{L}_n(\widehat{\theta}_n) \equiv M$
- Algorithm
    1. Draw $\theta^{cand} \sim f(\theta)$
    2. Generate $u \sim \text{Unif}\,(0,1)$
    3. Accept θ^{cand} if $u \le \dfrac{\mathcal{L}_n(\theta^{cand})}{\mathcal{L}_n(\widehat{\theta}_n)}$

16.4 Importance Sampling

Sample from an importance function g rather than target density h.
Algorithm to obtain an approximation to $\mathbb{E}[q(\theta) \mid x^n]$:

1. Sample from the prior $\theta_1, \dots, \theta_n \overset{iid}{\sim} f(\theta)$
2. $w_i = \dfrac{\mathcal{L}_n(\theta_i)}{\sum_{i=1}^B \mathcal{L}_n(\theta_i)} \quad \forall i = 1, \dots, B$
3. $\mathbb{E}[q(\theta) \mid x^n] \approx \sum_{i=1}^B q(\theta_i) w_i$

17 Decision Theory

Definitions

- Unknown quantity affecting our decision: $\theta \in \Theta$
- Decision rule: synonymous for an estimator $\widehat{\theta}$
- Action $a \in \mathcal{A}$: possible value of the decision rule. In the estimation context, the action is just an estimate of θ, $\widehat{\theta}(x)$.
- Loss function L: consequences of taking action a when true state is θ or discrepancy between θ and $\widehat{\theta}$, $L : \Theta \times \mathcal{A} \to [-k, \infty)$.

Loss functions

- Squared error loss: $L(\theta, a) = (\theta - a)^2$
- Linear loss: $L(\theta, a) = \begin{cases} K_1(\theta - a) & a - \theta < 0 \\ K_2(a - \theta) & a - \theta \ge 0 \end{cases}$
- Absolute error loss: $L(\theta, a) = |\theta - a| \quad$ (linear loss with $K_1 = K_2$)
- L_p loss: $L(\theta, a) = |\theta - a|^p$
- Zero-one loss: $L(\theta, a) = \begin{cases} 0 & a = \theta \\ 1 & a \ne \theta \end{cases}$

17.1 Risk

Posterior risk

$$r(\widehat{\theta} \mid x) = \int L(\theta, \widehat{\theta}(x)) f(\theta \mid x)\, d\theta = \mathbb{E}_{\theta|X}\left[L(\theta, \widehat{\theta}(x))\right]$$

(Frequentist) risk

$$R(\theta, \widehat{\theta}) = \int L(\theta, \widehat{\theta}(x)) f(x \mid \theta)\, dx = \mathbb{E}_{X|\theta}\left[L(\theta, \widehat{\theta}(X))\right]$$

Bayes risk

$$r(f,\widehat{\theta}) = \iint L(\theta,\widehat{\theta}(x))f(x,\theta)\,dx\,d\theta = \mathbb{E}_{\theta,X}\left[L(\theta,\widehat{\theta}(X))\right]$$

$$r(f,\widehat{\theta}) = \mathbb{E}_{\theta}\left[\mathbb{E}_{X|\theta}\left[L(\theta,\widehat{\theta}(X))\right]\right] = \mathbb{E}_{\theta}\left[R(\theta,\widehat{\theta})\right]$$

$$r(f,\widehat{\theta}) = \mathbb{E}_{X}\left[\mathbb{E}_{\theta|X}\left[L(\theta,\widehat{\theta}(X))\right]\right] = \mathbb{E}_{X}\left[r(\widehat{\theta}\,|\,X)\right]$$

17.2 Admissibility

- $\widehat{\theta}'$ dominates $\widehat{\theta}$ if

$$\forall\theta : R(\theta,\widehat{\theta}') \le R(\theta,\widehat{\theta})$$

$$\exists\theta : R(\theta,\widehat{\theta}') < R(\theta,\widehat{\theta})$$

- $\widehat{\theta}$ is inadmissible if there is at least one other estimator $\widehat{\theta}'$ that dominates it. Otherwise it is called admissible.

17.3 Bayes Rule

Bayes rule (or Bayes estimator)

- $r(f,\widehat{\theta}) = \inf_{\widetilde{\theta}} r(f,\widetilde{\theta})$
- $\widehat{\theta}(x) = \inf r(\widehat{\theta}\,|\,x)\;\forall x \implies r(f,\widehat{\theta}) = \int r(\widehat{\theta}\,|\,x)f(x)\,dx$

Theorems

- Squared error loss: posterior mean
- Absolute error loss: posterior median
- Zero-one loss: posterior mode

17.4 Minimax Rules

Maximum risk

$$\bar{R}(\widehat{\theta}) = \sup_{\theta} R(\theta,\widehat{\theta}) \qquad \bar{R}(a) = \sup_{\theta} R(\theta,a)$$

Minimax rule

$$\sup_{\theta} R(\theta,\widehat{\theta}) = \inf_{\widetilde{\theta}} \bar{R}(\widetilde{\theta}) = \inf_{\widetilde{\theta}} \sup_{\theta} R(\theta,\widetilde{\theta})$$

$$\widehat{\theta} = \text{Bayes rule} \;\wedge\; \exists c : R(\theta,\widehat{\theta}) = c$$

Least favorable prior

$$\widehat{\theta}^f = \text{Bayes rule} \;\wedge\; R(\theta,\widehat{\theta}^f) \le r(f,\widehat{\theta}^f)\;\forall\theta$$

18 Linear Regression

Definitions

- Response variable Y
- Covariate X (aka predictor variable or feature)

18.1 Simple Linear Regression

Model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad \mathbb{E}\left[\epsilon_i\,|\,X_i\right] = 0,\; \mathbb{V}\left[\epsilon_i\,|\,X_i\right] = \sigma^2$$

Fitted line

$$\widehat{r}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

Predicted (fitted) values

$$\widehat{Y}_i = \widehat{r}(X_i)$$

Residuals

$$\hat{\epsilon}_i = Y_i - \widehat{Y}_i = Y_i - \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_i\right)$$

Residual sums of squares (RSS)

$$\textsc{rss}(\widehat{\beta}_0,\widehat{\beta}_1) = \sum_{i=1}^{n}\hat{\epsilon}_i^2$$

Least square estimates

$$\widehat{\beta}^T = (\widehat{\beta}_0,\widehat{\beta}_1)^T : \min_{\widehat{\beta}_0,\widehat{\beta}_1} \textsc{rss}$$

$$\widehat{\beta}_0 = \bar{Y}_n - \widehat{\beta}_1\bar{X}_n$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (X_i-\bar{X}_n)(Y_i-\bar{Y}_n)}{\sum_{i=1}^n (X_i-\bar{X}_n)^2} = \frac{\sum_{i=1}^n X_iY_i - n\bar{X}\bar{Y}}{\sum_{i=1}^n X_i^2 - n\overline{X^2}}$$

$$\mathbb{E}\left[\widehat{\beta}\,|\,X^n\right] = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$$

$$\mathbb{V}\left[\widehat{\beta}\,|\,X^n\right] = \frac{\sigma^2}{ns_X^2}\begin{pmatrix} n^{-1}\sum_{i=1}^n X_i^2 & -\overline{X}_n \\ -\overline{X}_n & 1\end{pmatrix}$$

$$\widehat{\mathsf{se}}(\widehat{\beta}_0) = \frac{\widehat{\sigma}}{s_X\sqrt{n}}\sqrt{\frac{\sum_{i=1}^n X_i^2}{n}}$$

$$\widehat{\mathsf{se}}(\widehat{\beta}_1) = \frac{\widehat{\sigma}}{s_X\sqrt{n}}$$

where $s_X^2 = n^{-1}\sum_{i=1}^n (X_i-\overline{X}_n)^2$ and $\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n \hat{\epsilon}_i^2$ (unbiased estimate). Further properties:

- Consistency: $\widehat{\beta}_0 \xrightarrow{\text{P}} \beta_0$ and $\widehat{\beta}_1 \xrightarrow{\text{P}} \beta_1$