Discrete likelihood		
Likelihood	Conjugate prior	Posterior hyperparameters
$\text{Bern}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+\sum_{i=1}^n x_i, \beta+n-\sum_{i=1}^n x_i$
$\text{Bin}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+\sum_{i=1}^n x_i, \beta+\sum_{i=1}^n N_i-\sum_{i=1}^n x_i$
$\text{NBin}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+rn, \beta+\sum_{i=1}^n x_i$
$\text{Po}(\lambda)$	$\text{Gamma}(\alpha,\beta)$	$\alpha+\sum_{i=1}^n x_i, \beta+n$
$\text{Multinomial}(p)$	$\text{Dir}(\alpha)$	$\alpha+\sum_{i=1}^n x^{(i)}$
$\text{Geo}(p)$	$\text{Beta}(\alpha,\beta)$	$\alpha+n, \beta+\sum_{i=1}^n x_i$

15.4 Bayesian Testing

If $H_0 : \theta \in \Theta_0$:

$$\text{Prior probability } \mathbb{P}[H_0] = \int_{\Theta_0} f(\theta)\, d\theta$$

$$\text{Posterior probability } \mathbb{P}[H_0 \mid x^n] = \int_{\Theta_0} f(\theta \mid x^n)\, d\theta$$

Let $H_0 \ldots H_{k-1}$ be k hypotheses. Suppose $\theta \sim f(\theta \mid H_k)$,

$$\mathbb{P}[H_k \mid x^n] = \frac{f(x^n \mid H_k)\mathbb{P}[H_k]}{\sum_{k=1}^K f(x^n \mid H_k)\mathbb{P}[H_k]},$$

Marginal likelihood

$$f(x^n \mid H_i) = \int_{\Theta} f(x^n \mid \theta, H_i) f(\theta \mid H_i)\, d\theta$$

Posterior odds (of H_i relative to H_j)

$$\frac{\mathbb{P}[H_i \mid x^n]}{\mathbb{P}[H_j \mid x^n]} = \underbrace{\frac{f(x^n \mid H_i)}{f(x^n \mid H_j)}}_{\text{Bayes Factor } BF_{ij}} \times \underbrace{\frac{\mathbb{P}[H_i]}{\mathbb{P}[H_j]}}_{\text{prior odds}}$$

Bayes factor

$\log_{10} BF_{10}$	BF_{10}	evidence
$0-0.5$	$1-1.5$	Weak
$0.5-1$	$1.5-10$	Moderate
$1-2$	$10-100$	Strong
>2	>100	Decisive

$$p^* = \frac{\frac{p}{1-p} BF_{10}}{1+\frac{p}{1-p} BF_{10}} \text{ where } p = \mathbb{P}[H_1] \text{ and } p^* = \mathbb{P}[H_1 \mid x^n]$$

16 Sampling Methods

16.1 Inverse Transform Sampling

Setup

- $U \sim \text{Unif}(0,1)$
- $X \sim F$
- $F^{-1}(u) = \inf\{x \mid F(x) \geq u\}$

Algorithm

1. Generate $u \sim \text{Unif}(0,1)$
2. Compute $x = F^{-1}(u)$

16.2 The Bootstrap

Let $T_n = g(X_1, \ldots, X_n)$ be a statistic.

1. Estimate $\mathbb{V}_F[T_n]$ with $\mathbb{V}_{\widehat{F}_n}[T_n]$.
2. Approximate $\mathbb{V}_{\widehat{F}_n}[T_n]$ using simulation:
   (a) Repeat the following B times to get $T^*_{n,1}, \ldots, T^*_{n,B}$, an IID sample from the sampling distribution implied by $\widehat{F}_n$
      i. Sample uniformly $X^*_1, \ldots, X^*_n \sim \widehat{F}_n$.
      ii. Compute $T^*_n = g(X^*_1, \ldots, X^*_n)$.
   (b) Then

$$v_{boot} = \widehat{\mathbb{V}}_{\widehat{F}_n} = \frac{1}{B}\sum_{b=1}^B \left(T^*_{n,b} - \frac{1}{B}\sum_{r=1}^B T^*_{n,r}\right)^2$$

16.2.1 Bootstrap Confidence Intervals

Normal-based interval

$$T_n \pm z_{\alpha/2}\widehat{\mathsf{se}}_{boot}$$

Pivotal interval

1. Location parameter $\theta = T(F)$

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 $k(\theta) \le Mg(\theta) \quad \forall\theta$

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

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