

1. Introduction

- By Central Limit Theorem, if sample (x_1, \dots, x_n) with each $X_i \sim D(\mu, \sigma^2)$ (D is some distribution) then as $n \rightarrow \infty$,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

So distribution of sample mean always tends to normal distribution, with standard deviation σ / \sqrt{n} .

- Unbiased estimate of standard deviation of sample mean:**

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Standard error of sample mean:** estimate of standard deviation of sample mean: s / \sqrt{n} .
- If n too small then s is poor estimator and mean may not be normally distributed.
- If population distribution is normal and n small then sample mean is t -distributed:

$$\frac{X - \mu}{s / \sqrt{n}} \sim t_{n-1}$$

$\frac{X - \mu}{s / \sqrt{n}}$ is **pivotal quantity** as distribution doesn't depend on parameters of X .

- Hypothesis test** for \underline{x} :
 - Define **null hypothesis** which identifies distribution believed to have generated each x_i .
 - Choose **test statistic** h (function of \underline{x}), extreme when null is false, not extreme when null is true.
 - Observed test statistic** is $t = h(\underline{x})$.
 - Determine how extreme t is as a realisation of $T = h(X_1, \dots, X_N)$ (so need to know distribution of T).
- One sided p -value:**

$$\mathbb{P}(T \geq t \mid H_0 \text{ true}) \quad \text{or} \quad \mathbb{P}(T \leq t \mid H_0 \text{ true})$$

- Two sided p -value:**

$$\mathbb{P}(T \geq |t| \cup T \leq -|t| \mid H_0 \text{ true})$$

2. Monte Carlo testing

- Monte Carlo testing:** given observed test stat $t = h(\underline{x})$, distribution $F(x \mid \theta)$, hypotheses $H_0 : \theta = \theta_0$, $H_1 : \theta > \theta_0$:
 - For $j \in \{1, \dots, N\}$:
 - Simulate n observations (z_1, \dots, z_n) from $F(\cdot \mid \theta_0)$.
 - Compute $t_j = h(z_1, \dots, z_n)$.
 - Estimate p -value by

$$P(T \geq t \mid H_0 \text{ true}) \approx \hat{p} = \frac{1}{N} \sum_{j=1}^N \mathbb{I}\{t_j \geq t\}$$

- **Resampling risk:** probability that Monte Carlo simulated p -value and true p -value are on different sides of significance threshold α (situation where Monte Carlo test is incorrect):

$$\text{resampling risk} = \begin{cases} \mathbb{P}(\hat{p} > \alpha) & \text{if } p \leq \alpha \\ \mathbb{P}(\hat{p} \leq \alpha) & \text{if } p > \alpha \end{cases}$$

3. The bootstrap

- **The non-parametric bootstrap estimate:** given independent data $\underline{x} = (x_1, \dots, x_n)$ and stat $S(\cdot)$, **resample** (draw samples of size n with replacement) \underline{x} B times to give $\underline{x}^{*1}, \dots, \underline{x}^{*B}$. To compute **bootstrap estimate of standard error of S** , compute

$$\widehat{\text{Var}}(S(\underline{x})) = \frac{1}{B-1} \sum_{b=1}^B (S(\underline{x}^{*b}) - \bar{S}^*)^2$$

where

$$\bar{S}^* = \frac{1}{B} \sum_{b=1}^B S(\underline{x}^{*b})$$

The standard error estimate is then $\sqrt{\widehat{\text{Var}}(S(\underline{x}))}$, i.e. the standard deviation of $S(\underline{x}^{*1}), \dots, S(\underline{x}^{*B})$. The **bootstrap estimate** of S is simply $S(\underline{x})$.

- For random variable X , **(cumulative) distribution function (cdf)** $F : \mathbb{R} \rightarrow [0, 1]$ is

$$F_X(x) = F(x) := \mathbb{P}(X \leq x)$$

- Properties of cdf:
 - $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.
 - **Monotonicity:** $x' < x \implies F(x') \leq F(x)$.
 - **Right-continuity:** $\lim_{t \rightarrow x^+} F(t) = F(x)$.
- Given data (x_1, \dots, x_n) with each sample i.i.d. realisation of random variable X , **empirical (cumulative) distribution function (ecdf)** is

$$\hat{F}(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{x_i \leq x\}$$

- **Glivenko-Cantelli theorem:** Let X_1, \dots, X_n be random sample from distribution with cdf F . Then

$$\sup_{x \in \mathbb{R}} |\hat{F}(x) - F(x)| \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

- Given data (x_1, \dots, x_n) , sampling uniformly at random from \underline{x} is equivalent to sampling from distribution with cdf defined as ecdf constructed from \underline{x} .
- For mean of sample of m draws from ecdf constructed from n data points, expectation and variance are

$$\mathbb{E}[\bar{Y}] = \bar{x}, \quad \text{Var}(\bar{Y}) = \frac{n-1}{n} \frac{s_x^2}{m}$$

- If S is the mean, $\widehat{\text{Var}}(S(\underline{x})) \rightarrow \frac{n-1}{n} \frac{s^2}{n}$ as $B \rightarrow \infty$.
- If **sampling fraction** $f = \frac{n}{N}$ where N population size, n sample size, is $f \geq 0.1$, can't assume infinite population.
- Given finite population of size N , mean \bar{X} of sample drawn uniformly at random without replacement has variance

$$\text{Var}(\bar{X}) = \frac{N-n}{N-1} \frac{\sigma^2}{n}$$

where σ^2 is true population variance.

- Given finite population of size N , sample of size n with variance S^2 drawn without replacement,

$$\mathbb{E} \left[\left(1 - \frac{n}{N} \right) \frac{S^2}{n} \right] = \text{Var}(\bar{X})$$

so it is unbiased estimator of $\text{Var}(\bar{X})$

- **Population bootstrap:** given independent data (x_1, \dots, x_n) drawn from finite population of size N , assuming $N/n = k$ is integer, construct new data set

$$\tilde{\underline{x}} = (x_1, \dots, x_n, x_1, \dots, x_n, \dots, x_1, \dots, x_n)$$

by repeating \underline{x} k times. Then construct B new samples $\underline{x}^{*1}, \dots, \underline{x}^{*B}$ by sampling without replacement. Then compute

$$\widehat{\text{Var}}(S(\underline{x})) = \frac{1}{B-1} \sum_{b=1}^B (S(\underline{x}^{*b}) - \bar{S}^*)^2$$

where

$$\bar{S}^* = \frac{1}{B} \sum_{b=1}^B S(\underline{x}^{*b})$$

If N/n not integer, $N = kn + m$ for $0 < m < n$, then before each of the B samples, append to $\tilde{\underline{x}}$ a sample without replacement of size m from \underline{x} .

- If data believed to follow type of distribution, can use **parametric bootstrap:** given independent data (x_1, \dots, x_n) , believed to be drawn from distribution $F(\cdot, \theta)$ with parameter θ :
 - Find maximum likelihood estimator $\hat{\theta}$.
 - Draw B new samples of size n from $F(\cdot, \hat{\theta})$ to give $\underline{x}^{*1}, \dots, \underline{x}^{*B}$.
 - Compute

$$\widehat{\text{Var}}(S(\underline{x})) = \frac{1}{B-1} \sum_{b=1}^B (S(\underline{x}^{*b}) - \bar{S}^*)^2$$

where

$$\bar{S}^* = \frac{1}{B} \sum_{b=1}^B S(\underline{x}^{*b})$$

- For parameter θ of distribution, estimated by statistic S , with $\hat{\theta} = S(\underline{x})$, **bias** is

$$\text{bias}(\theta, \hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

- **Basic bootstrap bias estimate:**

$$\widehat{\text{bias}}(\theta, \hat{\theta}) = \bar{S}^* - \hat{\theta} = \frac{1}{B} \sum_{b=1}^B S(\underline{x}^{*b}) - S(\underline{x})$$

- **Bias correction:** subtract bias from usual estimate:

$$\hat{\theta} - \widehat{\text{bias}}(\theta, \hat{\theta}) = 2\hat{\theta} - \bar{S}^*$$

But often $2\hat{\theta} - \bar{S}^*$ has higher variance as estimator than $\hat{\theta}$.

- **Normal confidence interval for bootstrap estimate:** $100(1 - \alpha)\%$ confidence interval is

$$\hat{\theta} \pm z_{\alpha/2} \sqrt{\widehat{\text{Var}}(S(\underline{x}))}$$

where $z_{\alpha/2}$ is $100(\alpha / 2)\%$ percentile of standard normal distribution. **Note:** only valid if size of data large enough, need to check for normality of bootstrap samples using quantile plot.

- **Percentile confidence interval:** use if \hat{F} close to true distribution. $100(1 - \alpha)\%$ confidence interval is

$$[S_{((\alpha/2)B)}^*, S_{((1-\alpha/2)B)}^*]$$

where $S_{(i)}^*$ is i th largest value of $S(\underline{x}^{*b})$ for $b = 1, \dots, B$. B must be chosen to make $(\alpha / 2)B$ and $(1 - \alpha / 2)B$ integers. B must be > 2000 for this to be good estimate. **Note:** inaccurate if bias or non-constant standard error or distribution of $S(X) \mid \theta$ isn't symmetric.

- **BC (bias corrected)** and **BCa (bias corrected and accelerated)** confidence intervals make adjustments when bias is present or there is non-constant standard error.

4. Monte Carlo integration

- Let random variable Y take values in sample space Ω with pdf f_Y , then

$$\mu := \mathbb{E}[Y] = \int_{\Omega} y f_Y(y) dy$$

- μ approximated by

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

for i.i.d. samples Y_i .

- If $Y = g(X)$ with X random variable with pdf f_X , then

$$\mu = \mathbb{E}[Y] = \mathbb{E}[g(X)] = \int g(x) f_X(x) dx$$

- To estimate $\int_a^b f(x) dx$, use $X \sim \text{Unif}(a, b)$

$$\mu = \int_a^b f(x) dx = \int_a^b (b-a) f(x) \frac{1}{b-a} dx = \int_a^b (b-a) f(x) f_X(x) dx = \mathbb{E}[(b-a) f(X)]$$

which can be estimated by

$$\hat{\mu}_n = (b-a) \frac{1}{n} \sum_{i=1}^n f(X_i)$$

for i.i.d. samples X_i .

- If $\text{Var}(Y) = \sigma^2 < \infty$, Monte Carlo integration unbiased as $\mathbb{E}[\hat{\mu}_n] = \mu$.
- **Mean-square error:** $\text{Var}(\hat{\mu}_n) = \mathbb{E}[(\hat{\mu}_n - \mu)^2] = \frac{\sigma^2}{n}$.
- **Root mean-square error:** $\text{RMSE} = \sqrt{\mathbb{E}[(\hat{\mu}_n - \mu)^2]} = \frac{\sigma}{\sqrt{n}}$.
- RMSE is $O(n^{-1/2})$.
- For functions f, g , $f(n) = O(g(n))$ as $n \rightarrow \infty$ if exist $C, n_0 \in \mathbb{R}$ such that

$$\forall n \geq n_0, \quad |f(n)| \leq Cg(n)$$

- **Midpoint Riemann integral estimate:**

$$\int_a^b f(x) dx = \frac{b-a}{n} \sum_{i=1}^n f(x_i)$$

where

$$x_i = a + \frac{b-a}{n} \left(i - \frac{1}{2}\right)$$

- For d dimensions, Riemann sum converges in $O(n^{-2/d})$, Monte Carlo converges in $O(n^{-1/2})$ regardless of d .
- $100(1 - \alpha)\%$ confidence interval for Monte Carlo integration:

$$\mu \in \hat{\mu}_n \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where σ estimated with standard sample deviation of $\{y_i\} = \{g(x_i)\}$.

- If $g(x)$ constant multiple of indicator function, $g(x) = c\mathbb{I}\{A(x)\}$ for condition A , then

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{A(x_i)\}$$

is estimator for $p = \mathbb{P}(A)$. Binomial confidence interval is

$$p \in \hat{p}_n \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}}$$

so confidence interval for μ is

$$\mu \in \hat{\mu}_n \pm cz_{\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$$

$$(\hat{\mu}_n = c\hat{p}_n).$$

- Probability of no 1s in n Monte Carlo samples is $(1-p)^n$ so one-sided $100(1-\alpha)\%$ confidence interval has upper bound $p \leq 1 - \alpha^{1/n} \approx -\frac{\log(\alpha)}{n}$ using Taylor expansion.
- If \hat{p} very small and non-zero,

$$cz_{\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \approx cz_{\alpha/2} \sqrt{\frac{\hat{p}_n}{n}}$$

so relative error is

$$\delta := cz_{\alpha/2} \sqrt{\frac{\hat{p}_n}{n}} / \hat{p} = \frac{cz_{\alpha/2}}{\sqrt{\hat{p}_n n}}$$

for relative error at most δ ,

$$n \geq \frac{c^2 z_{\alpha/2}^2}{\hat{p}_n \delta^2}$$

so n grows inversely with \hat{p}_n .

- To estimate probability of event $\mathbb{P}(X \in E)$, Monte Carlo estimate $\mathbb{E}[\mathbb{I}\{X \in E\}]$.

5. Simulation

- Let F cdf, then **generalised inverse cdf** is

$$F^{-1}(u) := \inf\{x : F(x) \geq u\}$$

- **Inverse transform sampling algorithm:** let random variable X with cdf F , with generalised inverse F^{-1} .
 - Simulate $U \sim \text{Unif}(0, 1)$.
 - Compute $X = F^{-1}(U)$.

X is then distributed with cdf F . Only works for 1D distributions.

- **Rejection sampling algorithm:** given **target density** function f , **proposal density** function \tilde{f} with $\forall x \in \mathbb{R}^d, f(x) \leq c\tilde{f}(x)$ for some $c < \infty$,
 - Set $a = \text{false}$
 - While $a = \text{false}$:
 - Simulate $u \sim \text{Unif}(0, 1)$.
 - Simulate $x \sim \tilde{f}(\cdot)$.
 - If $u \leq \frac{f(x)}{c\tilde{f}(x)}$, set $a = \text{true}$.
 - Once while loop exited, return x , which is distributed with pdf f .
- **Note:** f and \tilde{f} don't need to be normalised.
- When f, \tilde{f} normalised, expected number of iterations of rejection sampling algorithm is c .
- **Important:** when choosing value of c , always round **up** if inexact.

- When checking if rejection sampling can be used, check if ratio $f(x) / \tilde{f}(x)$ tends to 0 as $x \rightarrow \pm\infty$ and differentiate ratio with respect to x to find maximum.
- **Normalised importance sampling:** given normalised density function f and normalised proposal density function \tilde{f} , n importance samples produced by: for $i \in \{1, \dots, n\}$:
 - Simulate $x_i \sim \tilde{f}(\cdot)$.
 - Compute $w_i = f(x_i) / \tilde{f}(x_i)$.

This produces importance samples $\{(x_i, w_i)\}_{i=1}^n$. $\mu = \mathbb{E}_{\tilde{f}}[g(X)]$ estimated by **importance sampling estimator**

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n w_i g(x_i)$$

$(\mathbb{E}_{\tilde{f}}[\hat{\mu}] = \mu, \text{ provided } \tilde{f}(x) > 0 \text{ whenever } f(x)g(x) \neq 0)$.

- Variance of importance sampling estimator is

$$\text{Var}(\hat{\mu}) = \frac{\sigma_{\tilde{f}}^2}{n}$$

where

$$\sigma_{\tilde{f}}^2 = \int_{\tilde{\Omega}} \frac{(g(x)f(x) - \mu\tilde{f}(x))^2}{\tilde{f}(x)} dx$$

and $\tilde{\Omega}$ is support of \tilde{f} .

- Can estimate variance with

$$\hat{\sigma}_{\tilde{f}}^2 = \frac{1}{n} \sum_{i=1}^n (w_i g(x_i) - \hat{\mu})^2$$

- Distribution which minimises estimator variance is

$$\tilde{f}_{\text{opt}}(x) = \frac{|g(x)|f(x)}{\int_{\Omega} |g(x)|f(x) dx}$$

- **Self-normalised importance sampling:** same as normalised importance sampling, but compute

$$\hat{\mu} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i g(x_i)$$

Can use for unnormalised density functions f, \tilde{f} . $\hat{\mu}$ is not unbiased.

- Approximate variance of self-normalised estimator:

$$\text{Var}(\hat{\mu}) \approx \frac{\sigma_{\tilde{f}}^2}{n}$$

where

$$\sigma_{\tilde{f}}^2 = \sum_{i=1}^n w_i'^2 (g(x_i) - \hat{\mu})^2$$

and

$$w_i' = \frac{w_i}{\sum_{j=1}^n w_j}$$

- **Effective sample size n_e** : size of sample for which variance of naive Monte Carlo average $\left(\frac{1}{n_e} \sum_{i=1}^{n_e} g(x_i)\right)$ with sample size n_e , σ^2 / n_e (σ^2 is variance of $g(X)$), is equal to variance of importance sampling estimator $\hat{\mu}$, $\text{Var}(\hat{\mu})$:

$$n_e = \frac{n\bar{w}^2}{\overline{w^2}}$$

where

$$\bar{w}^2 = \left(\frac{1}{n} \sum_{i=1}^n w_i\right)^2, \quad \overline{w^2} = \frac{1}{n} \sum_{i=1}^n w_i^2$$

- Small n_e means importance sampling is poor estimator.
- Poor estimator if proposal distribution has much less probability in tails than target distribution.