C1_Review of prob,Frequentist Statistics,The Bootstrap, and Random Variate Generation

1. Concept and Definition Review

1. **Variance**

$$\operatorname{var}(X)=E\left[\left(X-\mu_X\right)^2\right]=\int_{\mathbb{R}}\left(x-\mu_X\right)^2 dF(x)$$
$$\text{Note: We also denote the variance as } \sigma_X^2 \text{ .}$$

2. **Covariance**

$$\operatorname{cov}(X, Y)=E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]$$

   - **Correlation:**

$$\rho_{X, Y}=\operatorname{cov}(X, Y) /\left(\sigma_X \sigma_Y\right)$$

2. Common distributions

1. **Binomial random variable**, $X \sim$ Binomial (n, θ) if

$$f_X(k)=\mathbb{P}(X=k)=\binom{n}{k} \theta^k(1-\theta)^{n-k}, k=0,1, \ldots, n$$

2. **Geometric distribution** is the distribution of the number of Bernoulli trials until the first success.
   $f_X(n)=\mathbb{P}(X=n)=\theta(1-\theta)^{n-1}, n=1,2, \ldots$

3. **Poisson random variable**

$$\mathbb{P}(X=n)=e^{-\theta}\theta^n/n!, n=0,1,2,\ldots$$

4. **Normal distribution** denoted by $N\left(\mu, \sigma^2\right)$. The parameter μ is called the mean and the parameter σ^2 is called the variance.
    - probability density function $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$
    - If X is a normal random variable with mean μ and variance $\sigma^2>0$, then $Z=\frac{X-\mu}{\sigma}$ is a standard normal random variable.

5. **Gamma distribution** with shape parameter k and scale parameter θ, i.e., $X \sim \operatorname{Gamma}(k, \theta)$if it has PDF:
    - $f(x)=\frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-x/\theta}, x \geq 0$

3. Frequentist inference

1. Frequentist methods are based on **principle of repeated sampling**, include methods to determine:

- whether the data x are compatible with the model $\mathcal{M}$
- what conclusions can be drawn about a parameter θ.

2. **Point estimation** is a fundamental question in frequentist inference.

- Let the data $X_1, \ldots, X_n$ be iid with CDF F and $\theta=P(F)$ **is a scalar/vector of interest**. Then, a point estimate of θ, denoted by $\hat{\theta}$, is a single "best guess" of θ based on the data, i.e., $\hat{\theta}=T(X))$
- The **bias of the estimator $\hat{\theta}$ is bias** $(\hat{\theta})=E_F(\hat{\theta})-\theta$. We write E_F instead of E to emphasize that the expectation is with respect to F.
- The **mean squared error** $(\mathrm{MSE})$of $\hat{\theta}$ is $E_F\left[(\hat{\theta}-\theta)^2\right]=\operatorname{bias}^2(\hat{\theta})+\operatorname{var}(\hat{\theta})$
- A **point estimator** $\hat{\theta}$ of θ is consistent if $\hat{\theta} \rightarrow^P \theta$ as $n \rightarrow \infty$

3. **The empirical distribution function**
    - Definition 1.21: The **empirical distribution function** $\hat{F}_n$ is the CDF that puts probability mass $1/n$ at each data point X_i.:
        - $\hat{F}_n(x)=n^{-1}\sum_{i=1}^n I\left(X_i \leq x\right)$.
    - A **statistical function** $P(F)$ is any function of F.
        - Examples are $\mu=\int_{\mathbb{R}} x dF(x)$, the variance $\sigma^2=\int_{\mathbb{R}}(x-\mu)^2 dF(x)$, and the p th quantile $F^{-1}(p)=\inf\{x: F(x) \geq p\}$

- Definition 1.22: The **plug-in estimator** of $\theta = P(F)$ is defined by $\hat{\theta}_n = P\left(\widehat{F}_n\right) = T(X)$. In other words, **just replace F with $\hat{F}_n$**
  - Example: The plug-in estimator for a linear functional $P(F) = \int_{\mathbb{R}} r(x) dF(x)$ is $P\left(\widehat{F}_n\right) = \int_{\mathbb{R}} r(x) d\widehat{F}_n(x) = \left(n^{-1}\right) \sum_{i=1}^{n} r\left(X_i\right) = T(X)$
  - **Exercise example:**

    Since $\sigma_1^2 = E\left\{[X - E(X)]^2\right\} = E\left\{[X-1]^2\right\}$, and $\sigma_2^2 = E\left\{[Y - E(Y)]^2\right\} = E\left\{[Y-1]^2\right\}$, we have $\varphi = \sigma_1^2 \sigma_2^2 = E\left\{[X-1]^2\right\} E\left\{[Y-1]^2\right\}$.

    Thus, a plug-in estimator is $\hat{\varphi} = \left[\frac{1}{n} \sum_{i=1}^{n} \left(X_i - 1\right)^2\right] \left[\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - 1\right)^2\right]$.

4. The Bootstrap

1. **Aim**:

```
* estimating standard errors
* computing confidence intervals
```

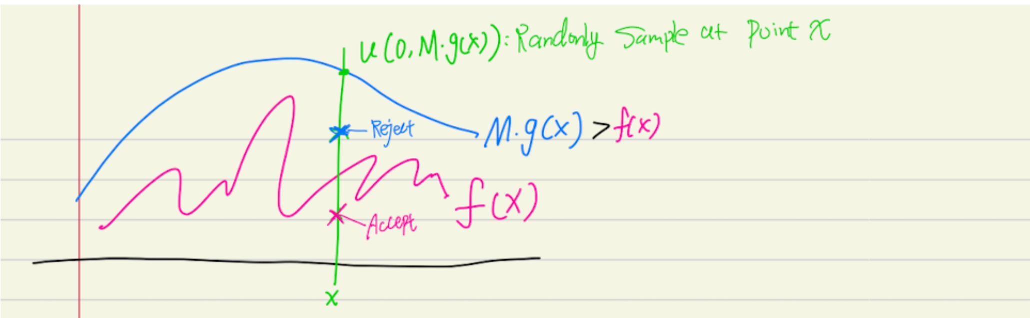
2. **2 steps:**

   Step 1: Estimate $\operatorname{var}_{X} \sim_F (T(X))$ with $\operatorname{var}_{X\sim} \hat{F}_n(T(X))$
   Step 2: Approximate $\operatorname{var}_{X \sim \hat{F}_n}(T(X))$ using Monte Carlo simulation

5. Random Variate Generation

1. Rejection Sampling Method

- intuition



> For: $U(0, M\cdot g(x)) \longrightarrow u'$

$u' > f(x)$ Reject

$u' < f(x)$ Accept

> And $u' = \underbrace{U(0,1)}_{u} * M\cdot g(x)$

$u \cdot M \cdot g(x) > f(x)$: Reject

$u \cdot M \cdot g(x) < f(x)$: Accept

So:

$\frac{f(x)}{M\cdot g(x)} > u$: Accept $\quad \frac{f(x)}{g(x)} > u\cdot M$

$\frac{f(x)}{M\cdot g(x)} < u$: Reject $\quad \frac{f(x)}{g(x)} < u\cdot M$

M: Maximum for $\frac{f(x)}{g(x)}$ find

$f(x)$: Source Pdf

$g(x)$: Proposal Pdf $\Longrightarrow$ generate Sample Y

- **Rejection sampling method concepts:**

  aim to generate function of f(x)

  - let g be a density function such that $f(x) \le Mg(x)$ for all x for some **constant** M. It is assumed that we can generate a random sample easily from g.
  - The rejection sampling method consists of the following steps:
    - 1. Generate $Y \sim g$ and $U \sim \text{unif}([0,1])$ independently
    - 2. If $U \le f(Y)/[Mg(Y)]$, set $X = Y$. Otherwise, return to 1.
  - Key: find the value of M

- **How to find M?**
  - Since $f(x) \le Mg(x)$, $f(x)/g(x) \le M$.
  - Find the **maximum point** for $f(x)/g(x)$ by take the **derivative == 0**

- solve the equation

- Example-1

  Derive an acceptance sampling scheme for generating from a **Gamma(k, θ) distribution** using the **exponential PDF** as g.

  - Note that $f(x)=\frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-x/\theta}, x\geq 0$ and $g(x)=\lambda e^{-\lambda x}, x\geq 0$
  - Define $r(x)=\frac{f(x)}{g(x)}=\frac{1}{\lambda\Gamma(k)\theta^k}x^{k-1}e^{(\lambda-1/\theta)x}, x\geq 0$
  - We find that $r'(x)=\frac{1}{\lambda\Gamma(k)\theta^k}x^{k-2}e^{(\lambda-1/\theta)x}\left[k-1+x\left(\lambda-\frac{1}{\theta}\right)\right]$
  - $r(x)$ achieves a maximum at $x^*=\frac{k-1}{\frac{1}{\theta}-\lambda}$
  - So, $M=r\left(x^*\right)=\frac{(k-1)^{k-1}}{\lambda\Gamma(k)\theta^k\left(\frac{1}{\theta}-\lambda\right)^{k-1}}e^{-(k-1)}$

- **Example-2**

Q1.

The probability density function of the beta distribution is

$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}I(0<x<1),$

where $\Gamma(\cdot)$ is the gamma function and $I(0<x<1)=\begin{cases}1, \text{if } 0<x<1\\ 0, \quad \text{otherwise}\end{cases}.$

Suppose that $a>1$ and $b>1$. Give a rejection sampling algorithm for sampling from this beta distribution using samples from the uniform distribution on $[0,1]$, i.e., $g(x)=I(0\leq x\leq 1)$.

Answer:

For $0<x<1$,

$\frac{d}{dx}f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}[(a-1)x^{a-2}(1-x)^{b-1}-(b-1)x^{a-1}(1-x)^{b-2}]$

$=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-2}(1-x)^{b-2}[(a-1)(1-x)-(b-1)x]$

$=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-2}(1-x)^{b-2}[(a-1)-(a+b-2)x].$

We see that there is a unique stationary point $x=x^*=\frac{a-1}{a+b-2}\in(0,1)$, $\frac{d}{dx}f(x)>0$ when $x<x^*$, and $\frac{d}{dx}f(x)<0$ when $x>x^*$. Thus, $x=x^*$ is the unique mode of f. Define

$M=f(x^*)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{a-1}{a+b-2}\right)^{a-1}\left(1-\frac{a-1}{a+b-2}\right)^{b-1}=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{a-1}{a+b-2}\right)^{a-1}\left(\frac{b-1}{a+b-2}\right)^{b-1}.$

It follows that $f(x)\leq Mg(x)$ for all x.

The rejection sampling algorithm is

1. Generate $Y\sim\text{unif}([0,1])$ and $U\sim\text{unif}([0,1])$ independently.
2. If $U\leq f(Y)/M$, set $X=Y$. Otherwise, return to 1.