Econ 488, First Problem Set

2025-01-16

Solutions:

1.

A subset of a population where the outcomes are independent and identically distributed (i.i.d.) between one sample and another.

2.

Yes, we expect each Y_i to be unaffected by others in terms of drawing outcomes and their chance of being drawn. For example, if $\{Y_1, Y_2\}$ or $\{Y_1, Y_{10}\}$ are drawn consecutively without clearing the draw box, they are still considered randomly sampled.

3.

Possible dependent and control variables for collecting December spending information:

- **Discrete:** Average spending during December.
- **Continuous:** Daily spending during December.

4.

```r
set.seed(1)

# Continuous random variable: N(mu = 5, sigma = 1)
N5_1 <- function(x) {
  x * dnorm(x, mean = 5, sd = 1)
}

EX <- integrate(N5_1, lower = -Inf, upper = Inf)
cat("The expected value of N(mu = 5, sigma = 1):", EX$value, "\n")
```

```
## The expected value of N(mu = 5, sigma = 1): 5
```

```r
# Discrete random variable: Bernoulli trials with outcomes 0 and 1
ber_values <- c(0, 1)
prob <- 1 / length(ber_values)
BT <- ber_values[1] * prob + ber_values[2] * prob
cat("The expected value of Bernoulli trials with outcomes 0, 1:", BT, "\n")
```

```
## The expected value of Bernoulli trials with outcomes 0, 1: 0.5
```

5.

i. Given $f(\cdot) = g(\mu; \theta)$, a probability density function (pdf) in parametric form, we deduce $\hat{\theta}$, a scalar, by solving for the parameters that maximize the likelihood function.

ii. Estimators are deduced based on random samples (Y_i), making the estimator itself random.

iii. An estimator is a formal function definition that approximates the true result, whereas an estimate is the empirical result computed using the estimator.

6.

An unbiased estimator satisfies $E[\hat{\theta}] - \theta = 0$.

- **Finite Sample Property:** When sample size increases, the estimator remains unbiased.
- **Bias and Variance:** A low-bias estimator does not guarantee low variance. Estimators may trade bias for variance depending on precision requirements.

Side notes: - **Monte Carlo Simulation (MCS):** Empirical distribution of $\hat{\theta}$ over simulated samples, accurate for large sample sizes. - **Central Limit Theorem (CLT):** Normal distribution approximation improves with larger samples.

7.

Given Y is i.i.d.:

1. $E[W_n^a] = E[Y_1]$.
2. $E[W_n^b] = E[Y_n]$.
3. $E[W_n^c] = E[(Y_n + Y_1)/2]$.
4. $E[W_n^d] = E[(Y_n + \cdots + Y_1)/n]$.

Thus, estimators a, b, c, and d are unbiased.

8.

A consistent estimator satisfies:

$$\lim_{T \to \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$$

Explanation: Consistency is a large-sample property where the probability of the estimator deviating from the true parameter diminishes as sample size increases.

9.

Law of Large Numbers (LLN):

$$\bar{X} \xrightarrow{P} \mu, \text{ as } n \to \infty$$

$$\lim_{n \to \infty} P(|\bar{X} - \mu| < \epsilon) = 1$$

10.

Central Limit Theorem (CLT):

$$\sqrt{T}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1), \text{ as } T \to \infty$$

11.

Law of Iterated Expectations (LIE):

$$E(X) = E_Y[E_X(X|Y)] = \sum_y P(Y=y)E(X|Y=y)$$

12.

```
PH1_T <- c(0.6, 0.2)  # P(T = 1 | H = 1) and P(T = 0 | H = 1)
PD1_H <- c(0.1, 0.5)  # P(D = 1 | H = 1) and P(D = 1 | H = 0)

# Probabilities for not deceased
PD0_H1 <- 1 - PD1_H[1]
PD0_H0 <- 1 - PD1_H[2]

# Total probabilities
PD0_T1 <- (PD0_H1 * PH1_T[1]) + (PD0_H0 * (1 - PH1_T[1]))
PD0_T0 <- (PD0_H1 * PH1_T[2]) + (PD0_H0 * (1 - PH1_T[2]))

cat("Traveled and not deceased:", PD0_T1 * 100, "%\n")
```

```
## Traveled and not deceased: 74 %
```

```
cat("Stayed home and not deceased:", PD0_T0 * 100, "%\n")
```

```
## Stayed home and not deceased: 58 %
```

13.

```
set.seed(25)
X <- 1:100
e <- rnorm(100, mean = 0, sd = 1)
Y <- 1 + 2 * X + e
Z <- lm(Y ~ X)

cat("Coefficients - Slope:", Z$coefficients[2], "Intercept:", Z$coefficients[1], "\n")
```

```
## Coefficients - Slope: 2.001709 Intercept: 0.7315357
```

```r
# Manually calculate slope
ls_slope <- cov(Y, X) / var(X)
cat("Manually calculated slope:", ls_slope, "\n")
```

```
## Manually calculated slope: 2.001709
```