Summary

Emanuel Mejía

Contents

1. Probability Theory

1.1 RANDOM EVENTS

Definition 1.1.1 ***Event Space***

A set S of subsets of Ω is an event space if it satisfies the following:

- Nonempty: $S \neq \emptyset$.
- Closed under complements: if $A \in S$, then $A^C \in S$.
- Closed under countable unions: if $A_1, A_2, A_3, ... \in S$, then $A_1 \cup A_2 \cup A_3 \cup ... \in S$.

Definition 1.1.2 ***Kolmogorov Axioms***

Let Ω be a sample space, S be an event space, and P be a probability measure. Then (Ω, S, P) is a probability space if it satisfies the following:

- Non-negativity: $\forall A \in S$, $P(A) \geq 0$, where $P(A)$ is finite and real.
- Unitarity: $P(\Omega) = 1$.
- Countable additivity: if $A1, A2, A3, ... \in S$ are pairwise disjoint, then

$$P(A_1 \cup A_2 \cup A_3 \cup ...) = P(A_1) + P(A_2) + P(A_3) + ... = \sum_i P(A_i)$$

Theorem 1.1.4 ***Basic Properties of Probability***

Let (Ω, S, P) be a probability space. Then:

- Monotonicity: $\forall A, B \in S$, if $A \subseteq B$, then $P(A) \leq P(B)$.
- Subtraction rule: $\forall A, B \in S$, if $A \subseteq B$, then $P(A \setminus B) = P(A) - P(B)$.
- Zero probability of the empty set: $P(\emptyset) = 0$.
- Probability bounds: $\forall A \in S$, $0 \leq P(A) \leq 1$.
- Complement rule: $\forall A \in S$, $P(A^C) = 1 - P(A)$

Definition 1.1.5. ***Joint Probability***

For $A, B \in S$, the joint probability of A and B is $P(A \cap B)$.

Theorem 1.1.7. ***Addition Rule***

For $A, B \in S$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definition 1.1.8. ***Conditional Probability***

For $A, B \in S$ with $P(B) > 0$, the conditional probability of A given B is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Theorem 1.1.9. ***Multiplicative Law of Probability***

For $A, B \in S$ with $P(B) > 0$, $P(A|B)P(B) = P(A \cap B)$.

Theorem 1.1.10. ***Bayes' Rule***

For $A, B \in S$ with $P(A) > 0$ and $P(B) > 0$,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Definition 1.1.12. ***Partition***

If $A_1, A_2, A_3, ... \in S$ are nonempty and pairwise disjoint, and $\Omega = A_1 \cup A_2 \cup A_3 \cup ...$, then $\{A_1, A_2, A_3, ...\}$ is a partition of Ω.

Theorem 1.1.13. ***Law of Total Probability***

If $\{A_1, A_2, A_3, ...\}$ is a partition of Ω and $B \in S$, then:

$$P(B) = \sum_i P(B \cap Ai)$$

If we also have $P(A_i) > 0$ for $i = 1, 2, 3, \ldots$, then this can also be stated as:

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

Theorem 1.1.14. ***Alternative Forms of Bayes' Rule***

If $\{A_1, A_2, A_3, ...\}$ is a partition of Ω with $P(A_i) > 0$ for $i = 1, 2, 3, \ldots$, and $B \in S$ with $P(B) > 0$, then:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B \cap A_i)}$$

or equivalently:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}$$

Definition 1.1.15. ***Independence of Events***

Events $A, B \in S$ are independent if $P(A \cap B) = P(A)P(B)$.

Theorem 1.1.16. ***Conditional Probability and Independence***

For $A, B \in S$ with $P(B) > 0$, A and B are independent if and only if $P(A|B) = P(A)$.

1.2 RANDOM VARIABLES

Definition 1.2.1. ***Random Variable***

A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that, $\forall r \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \leq r\} \in S$.

Definition 1.2.2. ***Function of a Random Variable***

Let $g : U \to \mathbb{R}$ be a function, where $X(\Omega) \subseteq U \subseteq \mathbb{R}$. Then, if $(g \circ X) : \Omega \to \mathbb{R}$ is a random variable, we say that g is a function of X, and write $g(X)$ to denote the random variable $(g \circ X)$.

Definition 1.2.3. ***Operator on a Random Variable***

An operator A on a random variable maps the function $X(\cdot)$ to a real number, denoted by $A[X]$.

Definition 1.2.4. ***Discrete Random Variable***

A random variable X is discrete if its range, $X(\Omega)$, is a countable set.

Definition 1.2.5. ***Probability Mass Function (PMF)***

For a discrete random variable X, the probability mass function of X is $f(x) = Pr[X = x], \forall x \in \mathbb{R}$

Theorem 1.2.9. ***Properties of PMFs***

For a discrete random variable X with PMF f,

- $\forall x \in \mathbb{R}, f(x) \geq 0$.
- $\sum_{x \in X(\Omega)} f(x) = 1$.

Theorem 1.2.10. ***Event Probabilities for Discrete Random Variables***

For a discrete random variable X with PMF f, if $D \subseteq R$ and $A = \{X \in D\}$, then

$$P(A) = Pr\left[X \in D\right] = \sum_{x \in X(A)} f(x)$$

Definition 1.2.11. ***Cumulative Distribution Function (CDF)***

For a random variable X, the cumulative distribution function of X is

$$F(x) = Pr\left[X \leq x\right], \forall x \in R$$

Theorem 1.2.12. ***Properties of CDFs***

For a random variable X with CDF F,

- F is nondecreasing: $\forall x_1, x_2 \in \mathbb{R}$, if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $\forall x \in \mathbb{R}$, $1 - F(x) = Pr\left[X > x\right]$

Definition 1.2.14. ***Continuous Random Variable***

A random variable X is continuous if there exists a non-negative function $f : \mathbb{R} \to \mathbb{R}$ such that the CDF of X is:

$$F(x) = Pr\left[X \leq x\right] = \int_{-\infty}^{x} f(u)du, \forall x \in \mathbb{R}$$

Definition 1.2.15. ***Probability Density Function (PDF)***

For a continuous random variable X with CDF F, the probability density function of X is:

$$f(x) = \left.\frac{dF(u)}{du}\right|_{u=x}, \forall x \in \mathbb{R}$$

Theorem 1.2.16. ***Properties of PDFs***

For a continuous random variable X with PDF f,

- $\forall x \in \mathbb{R}, f(x) \geq 0$.
- $\int_{-\infty}^{\infty} f(x)dx = 1$.

Theorem 1.2.17. ***Event Probabilities for Continuous Random Variables***

For a continuous random variable X with PDF f,

- $\forall x \in \mathbb{R}, Pr[X = x] = 0$.
- $\forall x \in \mathbb{R}, Pr[X < x] = Pr[X \leq x] = F(x) = \int_{-\infty}^{x} f(u)du$.
- $\forall x \in \mathbb{R}, Pr[X > x] = Pr[X \geq x] = 1 - F(x) = \int_{x}^{\infty} f(u)du$.
- $\forall a, b \in \mathbb{R}$ with $a \leq b$,

$$\begin{aligned} Pr[a < X < b] = Pr[a \leq X < b] = Pr[a < X \leq b] = Pr[a \leq X \leq b] \\ = F(b) - F(a) = \int_{a}^{b} f(x)dx \end{aligned}$$

Definition 1.2.20. ***Support***

For a random variable X with PMF/PDF f, the support of X is:

$$Supp[X] = \{x \in \mathbb{R} : f(x) > 0\}$$

1.3 BIVARIATE RELATIONSHIPS

Definition 1.3.1. ***Equality of Random Variables***

Let X and Y be random variables. Then $X = Y$ if, $\forall \omega \in \Omega, X(\omega) = Y(\omega)$.

Theorem 1.3.2. ***Equality of Functions of a Random Variable***

Let X be a random variable, and let f and g be functions of X. Then

$$g(X) = h(X) \iff \forall x \in X(\Omega), g(x) = h(x)$$

Definition 1.3.3. ***Joint PMF***

For discrete random variables X and Y, the joint PMF of X and Y is:

$$f(x,y) = Pr[X = x, Y = y], \forall x, y \in \mathbb{R}$$

Definition 1.3.4. ***Joint CDF***

For random variables X and Y, the joint CDF of X and Y is:

$$F(x,y) = Pr[X \leq x, Y \leq y], \forall x, y \in \mathbb{R}.$$

Theorem 1.3.6. ***Marginal PMF***

For discrete random variables X and Y with joint PMF f, the marginal PMF of Y is:

$$f_Y(y) = Pr[Y = y] = \sum_{x \in Supp[X]} f(x,y), \forall y \in R.$$

Definition 1.3.7. ***Conditional PMF***

For discrete random variables X and Y with joint PMF f, the conditional PMF of Y given $X = x$ is:

$$f_{Y|X}(y|x) = Pr[Y = y|X = x] = \frac{Pr[X = x, Y = y]}{Pr[X = x]} = \frac{f(x,y)}{f_X(x)}, \forall y \in \mathbb{R} \text{ and } \forall x \in Supp[X]$$

Theorem 1.3.9. ***Multiplicative Law for PMFs***

Let X and Y be two discrete random variables with joint PMF f. Then, $\forall x \in \mathbb{R}$ and $\forall y \in Supp[Y]$:

$$f_{X|Y}(x|y) f_Y(y) = f(x,y)$$

Definition 1.3.10. ***Jointly Continuous Random Variables***

Two random variables X and Y are jointly continuous if there exists a non-negative function $f : \mathbb{R}^2 \to \mathbb{R}$ such that the joint CDF of X and Y is:

$$F(x,y) = Pr[X \leq x, Y \leq y] = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v) dv du, \forall x, y \in \mathbb{R}.$$

Definition 1.3.11. ***Joint PDF***

For jointly continuous random variables X and Y with joint CDF F, the joint PDF of X and Y is:

$$f(x,y) = \left. \frac{\delta^2 F(u,v)}{\delta u \delta v} \right|_{u=x, v=y}, \forall x, y \in \mathbb{R}$$

Theorem 1.3.12. ***Event Probabilities for Bivariate Continuous Distributions***

For jointly continuous random variables X and Y with joint PDF f, if $D \subseteq \mathbb{R}^2$, then

$$Pr[(X,Y) \in D] = \iint_D f(x,y) dy dx$$

Theorem 1.3.13. ***Marginal PDF***

For jointly continuous random variables X and Y with joint PDF f, the marginal PDF of Y is:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx, \forall y \in \mathbb{R}.$$

Definition 1.3.14. ***Conditional PDF***

For jointly continuous random variables X and Y with joint PDF f, the conditional PDF of Y given $X = x$ is:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}, \forall y \in \mathbb{R} \text{ and } \forall x \in Supp[X].$$

Theorem 1.3.15. ***Multiplicative Law for PDFs***

Let X and Y be two jointly continuous random variables with joint PDF f. Then, $\forall x \in \mathbb{R}$ and $\forall y \in Supp[Y]$:

$$f_{X|Y}(x|y)f_Y(y) = f(x,y)$$

Definition 1.3.16. ***Independence of Random Variables***

Let X and Y be either two discrete random variables with joint PMF f or two jointly continuous random variables with joint PDF f. Then X and Y are independent if, $\forall x, y \in \mathbb{R}$,

$$f(x,y) = f_X(x)f_Y(y)$$

We write $X \perp\!\!\!\perp Y$ to denote that X and Y are independent

Theorem 1.3.17. ***Implications of Independence (Part I)***

Let X and Y be either two discrete random variables with joint PMF f or two jointly continuous random variables with joint PDF f. Then the following statements are equivalent (that is, each one implies all the others):

- $X \perp\!\!\!\perp Y$.
- $\forall x, y \in \mathbb{R}, f(x,y) = f_X(x)f_Y(y)$.
- $\forall x \in \mathbb{R}$ and $\forall y \in Supp[Y]$, $f_{X|Y}(x|y) = f_X(x)$.
- $\forall D, E \subseteq \mathbb{R}$, the events $\{X \in D\}$ and $\{Y \in E\}$ are independent.
- For all functions g of X and h of Y, $g(X) \perp\!\!\!\perp h(Y)$.

2. Summarizing Distributions

2.1 SUMMARY FEATURES OF RANDOM VARIABLES

Definition 2.1.1. ***Expected Value***

For a discrete random variable X with probability mass function (PMF) f, if $\sum_x |x| f(x) < \infty$, then the expected value of X is:

$$E[X] = \sum_x x f(x)$$

For a continuous random variable X with probability density function (PDF) f, if $\int_{-\infty}^{\infty} |x| f(x) < \infty$, then the expected value of X is:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Theorem 2.1.5. ***Expectation of a Function of a Random Variable (LOTUS)***

If X is a discrete random variable with PMF f and g is a function of X, then:

$$E[g(X)] = \sum_x g(x)f(x)$$

If X is a continuous random variable with PDF f and g is a function of X then:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Theorem 2.1.6. ***Properties of Expected Values***

For a random variable X, - $\forall c \in \mathbb{R}, E[c] = c$. - $\forall a \in \mathbb{R}, E[aX] = aE[X]$.

Definition 2.1.7. ***Expectation of a Bivariate Random Vector***

For a random vector (X, Y), the *expected value* of (X, Y) is:

$$E[(X, Y)] = (E[X], E[Y])$$

Theorem 2.1.8. ***Expectation of a Function of Two Random Variables***

For discrete random variables X and Y with joint PMF f, if h is a function of X and Y, then:

$$E[h(X, Y)] = \sum_x \sum_y h(x, y)f(x, y)$$

For jointly continuous random variables X and Y with joint PDF f, if h is a function of X and Y, then:

$$E[h(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)f(x, y)dydx$$

Theorem 2.1.9. ***Linearity of Expectations***

Let X and Y be random variables. Then, $\forall a, b, c \in \mathbb{R}$,

$$E[aX + bY + c] = aE[X] + bE[Y] + c$$

Definition 2.1.10. jth Raw Moment

For a random variable X and $j \in N$, the j^{th} raw moment of X is

$$\mu'_j = E[X^j]$$

Definition 2.1.11. ***jth Central Moment***

For a random variable X and $j \in N$, the j^{th} central moment of X is:

$$\mu_j = E[(X - E[X])^j]$$

Definition 2.1.12. ***Variance***

The variance of a random variable X is

$$V[X] = E[(X - E[X])^2]$$

Theorem 2.1.13. ***Alternative Formula for Variance***

For a random variable X,

$$V[X] = E[X^2] - E[X]^2$$

Theorem 2.1.14. ***Properties of Variance***

For a random variable X,

- $\forall c \in R, V[X + c] = V[X]$.
- $\forall a \in R, V[aX] = a^2 V[X]$.

Definition 2.1.15. ***Standard Deviation***

The standard deviation of a random variable X is:

$$\sigma[X] = \sqrt{V[X]}$$

Theorem 2.1.16. ***Properties of Standard Deviation***

For a random variable X,

- $\forall c \in \mathbb{R}, \sigma[X + c] = \sigma[X]$.
- $\forall a \in \mathbb{R}, \sigma[aX] = |a|\sigma[X]$.

Theorem 2.1.18. ***Chebyshev's Inequality***

Let X be a random variable with finite $\sigma[X] > 0$. Then, $\forall \varepsilon > 0$,

$$Pr\left[|X - E[X]| \geq \varepsilon\sigma[X]\right] \leq \frac{1}{\varepsilon^2}$$

Definition 2.1.19. ***Normal Distribution***

A continuous random variable X follows a *normal distribution* if it has PDF:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \forall x \in \mathbb{R}$$

for some constants $\mu, \sigma \in \mathbb{R}$ with $\sigma > 0$. We write $X \sim N(\mu, \sigma^2)$ to denote that X follows a normal distribution with parameters μ and σ.

Theorem 2.1.20. ***Mean and Standard Deviation of the Normal Distribution***

If $X \sim N(\mu, \sigma^2)$, then:

- $E[X] = \mu$.
- $\sigma[X] = \sigma$.

Theorem 2.1.21. ***Properties of the Normal Distribution***

Suppose $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Then:

- $\forall a, b \in \mathbb{R}$ with $a \neq 0$, if $W = aX + b$, then $W \sim N(a\mu_X + b, a^2\sigma_X^2)$.
- If $X \perp\!\!\!\perp Y$ and $Z = X + Y$, then $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Definition 2.1.22. ***Mean Squared Error (MSE) about c***

For a random variable X and $c \in \mathbb{R}$, the mean squared error of X about c is $E[(X - c)^2]$.

Theorem 2.1.23. ***Alternative Formula for MSE***

For a random variable X and $c \in \mathbb{R}$,

$$E\left[(X - c)^2\right] = V\left[X\right] + (E[X] - c)^2$$

Theorem 2.1.24. ***The Expected Value Minimizes MSE***

For a random variable X, the value of c that minimizes the MSE of X about c is $c = E[X]$.

2.2 SUMMARY FEATURES OF JOINT DISTRIBUTIONS

Definition 2.2.1. ***Covariance***

The covariance of two random variables X and Y is:

$$Cov[X, Y] = E\left[(X - E[X])(Y - E[Y])\right]$$

Theorem 2.2.2. ***Alternative Formula for Covariance***

For random variables X and Y

$$Cov[X, Y] = E[XY] - E[X]E[Y]$$

Theorem 2.2.3. ***Variance Rule***

Let X and Y be random variables. Then:

$$V[X + Y] = V[X] + 2Cov[X, Y] + V[Y]$$

More generally, $\forall a, b, c \in \mathbb{R}$,

$$V[aX + bY + c] = a^2V[X] + 2abCov[X, Y] + b^2V[Y]$$

Theorem 2.2.4. ***Properties of Covariance***

For random variables X, Y, Z, and W

- $\forall c, d \in \mathbb{R}, Cov[c, X] = Cov[X, c] = Cov[c, d] = 0$.
- $Cov[X, Y] = Cov[Y, X]$.
- $Cov[X, X] = V[X]$.
- $\forall a, b, c, d \in \mathbb{R}, Cov[aX + c, bY + d] = abCov[X, Y]$.
- $Cov[X + W, Y + Z] = Cov[X, Y] + Cov[X, Z] + Cov[W, Y] + Cov[W, Z]$.

Definition 2.2.5. ***Correlation***

The correlation of two random variables X and Y with $\sigma[X] > 0$ and $\sigma[Y] > 0$ is:

$$\rho[X, Y] = \frac{Cov[X, Y]}{\sigma[X]\sigma[Y]}$$

Theorem 2.2.6. ***Correlation and Linear Dependence***

For random variable X and Y,

- $\rho[X, Y] \in [-1, 1]$.
- $\rho[X, Y] = 1 \iff \exists a, b \in \mathbb{R}$ with $b > 0$ such that $Y = a + bX$.
- $\rho[X, Y] = -1 \iff \exists a, b \in \mathbb{R}$ with $b > 0$ such that $Y = a - bX$.

Theorem 2.2.7. ***Properties of Correlation***

For random variables X, Y, and Z,

- $\rho[X, Y] = \rho[Y, X]$.
- $\rho[X, X] = 1$.
- $\rho[aX + c, bY + d] = \rho[X, Y], \forall a, b, c, d \in \mathbb{R}$ such that either $a, b > 0$ or $a, b < 0$.
- $\rho[aX + c, bY + d] = -\rho[X, Y], \forall a, b, c, d \in \mathbb{R}$ such that either $a < 0 < b$ or $b < 0 < a$.

Theorem 2.2.8. ***Implications of Independence (Part II)***

If X and Y are independent random variables, then:

- $E[XY] = E[X]E[Y]$.
- Covariance is zero: $Cov[X, Y] = 0$.
- Correlation is zero: $\rho[X, Y] = 0$.
- Variances are additive: $V[X + Y] = V[X] + V[Y]$.

Definition 2.2.10. ***Conditional Expectation***

For discrete random variables X and Y with joint PMF f, the conditional expectation of Y given $X = x$ is:

$$E[Y|X = x] = \sum_y y f_{Y|X}(y|x), \text{ for } x \in Supp[X]$$

For jointly continuous random variables X and Y with joint PDF f, the conditional expectation of Y given $X = x$ is:

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy, \text{ for } x \in Supp[X]$$

Theorem 2.2.11. ***Conditional Expectation of a Function of Random***

Variables For discrete random variables X and Y with joint PMF f, if h is a function of X and Y, then the conditional expectation of $h(X,Y)$ given $X=x$ is:

$$E[h(X,Y)|X=x] = \sum_y h(x,y) f_{Y|X}(y|x), \text{ for } x \in Supp[X]$$

For jointly continuous random variables X and Y with joint PDF f, if h is a function of X and Y, then the conditional expectation of $h(X,Y)$ given $X=x$ is:

$$E[h(X,Y)|X=x] = \int_{-\infty}^{\infty} h(x,y) f_{Y|X}(y|x) dy, \text{ for } x \in Supp[X]$$

Definition 2.2.12. ***Conditional Variance***

For random variables X and Y, the conditional variance of Y given $X=x$ is:

$$V[Y|X=x] = E[(Y - E[Y|X=x])^2|X=x], \text{ for } x \in Supp[X]$$

Theorem 2.2.13. ***Alternative Formula for Conditional Variance***

For random variables X and Y, $\forall x \in Supp[X]$,

$$V[Y|X=x] = E[Y^2|X=x] - E[Y|X=x]^2$$

Theorem 2.2.14. ***Linearity of Conditional Expectations***

For random variables X and Y, if g and h are functions of X, then $\forall x \in Supp[X]$,

$$E[g(X)Y + h(X)|X=x] = g(x)E[Y|X=x] + h(x)$$

Definition 2.2.15. ***Conditional Expectation Function (CEF)***

For random variables X and Y with joint PMF/PDF f, the conditional expectation function of Y given $X=x$ is:

$$G_Y(x) = E[Y|X], \forall x \in Supp[X]$$

Theorem 2.2.17. ***Law of Iterated Expectations***

For random variables X and Y,

$$E[Y] = E_X[E[Y|X]]$$

Theorem 2.2.18. ***Law of Total Variance***

For random variables X and Y,

$$V[Y] = E[V[Y|X]] + V[E[Y|X]]$$

NOTE: First Term, Variance which is unexplainable by X. Second Term, Variance which is explainable by X

Theorem 2.2.19. ***Properties of Deviations from the CEF***

Let X and Y be random variables and let $\varepsilon = Y - E[Y|X]$. Then:

- $E[\varepsilon|X] = 0$.
- $E[\varepsilon] = 0$.
- If g is a function of X, then $Cov[g(X), \varepsilon] = 0$.
- $V[\varepsilon|X] = V[Y|X]$.
- $V[\varepsilon] = E[V[Y|X]]$.

Theorem 2.2.20. ***The CEF is the Best Predictor***

For random variables X and Y, the CEF, $E[Y|X]$, is the best (minimum MSE) predictor of Y given X.

Theorem 2.2.21. ***Best Linear Predictor (BLP)***

For random variables X and Y, if $V[X] > 0$, then the best (minimum MSE) linear predictor of Y given X is $g(X) = \alpha + \beta X$, where:

$$\alpha = E[Y] - \frac{Cov[X,Y]}{V[X]} E[X]$$

$$\beta = \frac{Cov[X,Y]}{V[X]}$$

Theorem 2.2.22. ***Properties of Deviations from the BLP***

Let X and Y be random variables and let $\varepsilon = Y - g(X)$, where $g(X)$ is the BLP. Then:

- $E[\varepsilon] = 0$.
- $E[X\varepsilon] = 0$.
- $Cov[X, \varepsilon] = 0$.

Theorem 2.2.25. ***Implications of Independence (Part III)***

If X and Y are independent random variables, then:

- $E[Y|X] = E[Y]$.
- $V[Y|X] = V[Y]$.
- The BLP of Y given X is $E[Y]$.
- If g is a function of X and h is a function of Y, then:
  - $E[g(Y)|h(X)] = E[g(Y)]$.
  - The BLP of $h(Y)$ given $g(X)$ is $E[h(Y)]$.

2.3 MULTIVARIATE GENERALIZATIONS

Definition 2.3.1. ***Covariance Matrix***

For a random vector X of length K, the covariance matrix $V[X]$ is a matrix whose (k,k)th entry is $Cov[X[k], X[k]]$, $\forall i, j \in \{1, 2, ..., K\}$. That is:

$$V[\mathbf{X}] = \begin{pmatrix} V[X_{[1]}] & Cov[X_{[1]}, X_{[2]}] & \cdots & Cov[X_{[1]}, X_{[K]}] \\ Cov[X_{[2]}, X_{[1]}] & V[X_{[2]}] & \cdots & Cov[X_{[2]}, X_{[K]}] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_{[K]}, X_{[1]}] & Cov[X_{[K]}, X_{[2]}] & \cdots & V[X_{[K]}] \end{pmatrix}$$

Theorem 2.3.2. ***Multivariate Variance Rule***

For random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$:

$$V[X_{[1]} + X_{[2]} + \cdots + X_{[K]}] = V\left[\sum_{k=1}^{K} X_{[k]}\right] = \sum_{k=1}^{K}\sum_{k'=1}^{K} Cov[X_{[k]}, X_{[k']}]$$

Definition 2.3.3. ***Conditional Expectation (Multivariate Case)***

For discrete random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y with joint PMF f, the conditional expectation of Y given $\mathbf{X} = \mathbf{x}$ is:

$$E[Y|\mathbf{X} = \mathbf{x}] = \sum_{y} y f_{Y|X}(y|x), \forall \mathbf{x} \in Supp[\mathbf{X}]$$

For jointly continuous random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y with joint PDF f, the conditional expectation of Y given $\mathbf{X} = \mathbf{x}$ is:

$$E[Y|\mathbf{X} = \mathbf{x}] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x), \forall \mathbf{x} \in Supp[\mathbf{X}]$$

Definition 2.3.4. ***CEF (Multivariate Case)***

For random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y with joint PMF/PDF f, the CEF of Y given $\mathbf{X} = \mathbf{x}$ is:

$$G_Y(\mathbf{x}) = E[Y|\mathbf{X} = \mathbf{x}], \forall \mathbf{x} \in Supp[\mathbf{X}]$$

Theorem 2.3.5. ***The CEF Is the Minimum MSE Predictor***

For random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y, the CEF, $E[Y|X]$, is the best (minimum MSE) predictor of Y given X.

Definition 2.3.6. ***BLP (Multivariate Case)***

For random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y, the best linear predictor of Y given X (that is, the minimum MSE predictor of Y given X among functions of the form $g(\mathbf{X}) = b_0 + b_1 X_{[1]} + b_2 X_{[2]} ... + b_K X_{[K]}$ is $g(\mathbf{X}) = \beta_0 + \beta_1 X_{[1]} + \beta_2 X_{[2]} ... + \beta_K X_{[K]}$), where

$$(\beta_0, \beta_1, \beta_2, ..., \beta_K) = argmin_{(b_0, b_1, b_2, ..., b_K) \in \mathbb{R}^{K+1}} E\left[\left(Y - \left(b_0 + b_1 X_{[1]} + b_2 X_{[2]} ... + b_K X_{[K]}\right)\right)^2\right]$$

Theorem 2.3.7. ***Coefficients of the BLP Are Partial Derivatives***

For random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y, if $g(\mathbf{X})$ is the best linear predictor of Y given $\mathbf{X}$, then $\forall k \in \{1, 2, ..., K\}$,

$$\beta_k = \frac{\delta g(X)}{\delta X_{[k]}}$$

Theorem 2.3.8. ***Properties of Deviations from the BLP (Multivariate Case)***

For random variables $X_{[1]}$, $X_{[2]}$, $\dots$, $X_{[K]}$, and Y, if $g(\mathbf{X})$ is the best linear predictor of Y given X and $\epsilon = Y - g(X)$, then:

- $E[\epsilon] = 0$.
- $\forall k \in \{1, 2, ..., K\}$, $E[X_{[k]}\epsilon] = 0$
- $\forall k \in \{1, 2, ..., K\}$, $Cov[X_{[k]}, \epsilon] = 0$.

Extra Theorem ***Coefficients of the BLP***

The BLP solution for the multivariate case is:

$$\beta = E[(\mathbf{X}^T\mathbf{X})^{-1}]E[\mathbf{X}^T Y]$$

3. Learning from Random Samples

3.1 I.I.D RANDOM VARIABLES

Definition 3.1.1. Independent and Identically Distributed (I.I.D.)

Let $X_1, X_2, ..., X_n$ be random variables with CDFs $F_1, F_2, ..., F_n$, respectively. Let F_A denote the joint CDF of the random variables with indices in the set A. Then $X_1, X_2, ..., X_n$ are independent and identically distributed if they satisfy the following:

- Mutually independent: $\forall A \subseteq \{1, 2, ..., n\}, \forall (x_1, x_2, ..., x_n) \in \mathbb{R}^n$,

  $$F_A\left((x_i)\right)_{i \in A} = \Pi_{i \in A} F_i(x_i)$$

- Identically distributed: $\forall i, j \in \{1, 2, ..., n\}$ and $\forall x \in \mathbb{R}$, $F_i(x) = F_j(x)$.

Definition 3.1.2. ***Finite Population Mass Function***

Given a finite population U with responses $x_1, x_2, ..., x_N$, the finite population mass function,

$$f_{FP}(x) = \frac{1}{N}\sum_{i=1}^{N} I(x_i = x)$$

.

3.2 ESTIMATION

Definition 3.2.1. ***Sample Statistic***

For i.i.d. random variables $X_1, X_2, ..., X_n$, a sample statistic is a function of $X_1, X_2, ..., X_n$:

$$T_{(n)} = h_{(n)}(X_1, X_2, ..., X_n)$$

where $h_{(n)} : \mathbb{R}^n \to \mathbb{R}, \forall n \in N$.

Definition 3.2.2. ***Sample Mean***

For i.i.d. random variables $X_1, X_2, ..., X_n$, the sample mean is:

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Theorem 3.2.3. ***The Expected Value of the Sample Mean Is the Population Mean***

For i.i.d. random variables $X_1, X_2, ..., X_n$,

$$E[\bar{X}] = E[X]$$

Theorem 3.2.4. ***Sampling Variance of the Sample Mean***

For i.i.d. random variables $X_1, X_2, ..., X_n$ with finite variance $V[X]$, the sampling variance of $\bar{X}$ is:

$$V[\bar{X}] = \frac{V[X]}{n}$$

Theorem 3.2.5. ***Chebyshev's Inequality for the Sample Mean***

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with finite variance $V[X] > 0$. Then, $\forall \epsilon > 0$,

$$Pr\left[|\bar{X} - E[X]| \geq \epsilon\right] \leq \frac{V[X]}{\epsilon^2 n}$$

Definition 3.2.6. ***Convergence in Probability***

Let $(T_{(1)}, T_{(2)}, T_{(3)}, ...)$ be a sequence of random variables and let $c \in \mathbb{R}$. Then $T_{(n)}$ converges in probability to c if, $\forall \epsilon > 0$,

$$\lim_{n\to\infty} Pr\left[|T_{(n)} - c| \geq \epsilon\right] = 0$$

or equivalently,

$$\lim_{n\to\infty} Pr\left[|T_{(n)} - c| < \epsilon\right] = 1$$

We write $T_{(n)} \xrightarrow{p} c$ to denote that $T_{(n)}$ converges in probability to c.

Theorem 3.2.7. ***Continuous Mapping Theorem (CMT)***

Let $(S_{(1)}, S_{(2)}, S_{(3)}, ...)$ and $(T_{(1)}, T_{(2)}, T_{(3)}, ...)$ be sequences of random variables. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a continuous function and let $a, b \in \mathbb{R}$. If $S_{(n)} \xrightarrow{p} a$ and $T_{(n)} \xrightarrow{p} b$, then

$$g(S_{(n)}, T_{(n)}) \xrightarrow{p} g(a, b)$$

Theorem 3.2.8. ***Weak Law of Large Numbers (WLLN)***

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with finite variance $V[X] > 0$ and let $\bar{X}_{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then:

$$\bar{X}_{(n)} \xrightarrow{p} E[X]$$

Theorem 3.2.9. ***Estimating the CDF***

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with common CDF F. Let $x \in \mathbb{R}$ and let $Z_i = I(X_i \leq x)$, $\forall i \in \{1, 2, ..., n\}$, where $I(\cdot)$ is the indicator function, that is, it takes the value one if its argument is true and zero if it is false. Then:

$$\bar{Z} \xrightarrow{p} F(x)$$

Definition 3.2.10. ***Unbiasedness***

An estimator $\hat{\theta}$ is unbiased for θ if $E[\hat{\theta}] = \theta$

Definition 3.2.11. ***Bias of an Estimator***

For an estimator $\hat{\theta}$, the bias of $\hat{\theta}$ in estimating θ is $E[\hat{\theta}] - \theta$.

Definition 3.2.12. ***Sampling Variance of an Estimator***

For an estimator $\hat{\theta}$, the sampling variance of $\hat{\theta}$ is $V[\hat{\theta}]$.

Definition 3.2.13. ***Standard Error of an Estimator***

For an estimator $\hat{\theta}$, the standard error of $\hat{\theta}$ is $\sigma[\hat{\theta}]$.

Definition 3.2.14. ***MSE of an Estimator***

For an estimator $\hat{\theta}$, the mean squared error (MSE) of $\hat{\theta}$ in estimating θ is $E[(\hat{\theta} - \theta)^2]$.

Theorem 3.2.15. ***Alternative Formula for the MSE of an Estimator***

For an estimator $\hat{\theta}$,$E[(\hat{\theta} - \theta)^2] = V[\hat{\theta}] + (E[\hat{\theta}] - \theta)^2$.

Definition 3.2.16. ***Relative Efficiency***

Let $\hat{\theta}_A$ and $\hat{\theta}_B$ be estimators of θ. Then $\hat{\theta}_A$ is more efficient than $\hat{\theta}_B$ if it has a lower MSE.

Definition 3.2.17. ***Consistency***

An estimator $\hat{\theta}$ is consistent for θ if $\hat{\theta} \xrightarrow{p} \theta$.

Definition 3.2.18. ***Plug-In Sample Variance***

For i.i.d. random variables $X_1, X_2, ..., X_n$, the plug-in sample variance is $\bar{X^2} - (\bar{X})^2$.

Theorem 3.2.19. ***Properties of the Plug-In Sample Variance***

For i.i.d. random variables $X_1, X_2, ..., X_n$ with finite variance $V[X] > 0$,

- $E[\bar{X^2} - (\bar{X})^2] = \frac{n-1}{n} V[X]$.
- $\bar{X^2} - (\bar{X})^2 \xrightarrow{p} V[X]$.

Definition 3.2.20. ***Unbiased Sample Variance***

For i.i.d. random variables $X_1, X_2, ..., X_n$, the unbiased sample variance is: $\hat{V}[X] = \frac{n}{n-1}(\bar{X^2} - (\bar{X})^2)$.

Theorem 3.2.21. ***Properties of the Unbiased Sample Variance***

For i.i.d. random variables $X_1, X_2, ..., X_n$ with finite variance $V[X] > 0$,

- $E\left[\hat{V}[X]\right] = V[X]$.
- $\hat{V}[X] \xrightarrow{p} V[X]$.

Definition 3.2.22. ***Convergence in Distribution***

Let $(T_{(1)}, T_{(2)}, T_{(3)}, ...)$ be a sequence of random variables with CDFs $(F_{(1)}, F_{(2)}, F_{(3)}, ...)$, and let T be a random variable with CDF F. Then $T_{(n)}$ converges in distribution to T if, $\forall t \in \mathbb{R}$ at which F is continuous,

$$\lim_{n \to \infty} F_{(n)}(t) = F(t)$$

We write $T_{(n)} \xrightarrow{d} T$ to denote that $T_{(n)}$ converges in distribution to T.

Definition 3.2.23. ***Standardized Sample Mean***

For i.i.d. random variables $X_1, X_2, ..., X_n$ with finite $E[X] = \mu$ and finite $V[X] = \sigma^2 > 0$, the standardized sample mean is:

$$Z = \frac{\left(X - E[\bar{X}]\right)}{\sigma[\bar{X}]} = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma}$$

Theorem 3.2.24. ***Central Limit Theorem***

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with (finite) $E[X] = \mu$ and finite $V[X] = \sigma^2 > 0$, and let Z be the standardized sample mean. Then

$$Z \xrightarrow{d} N(0, 1)$$

or equivalently,

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Theorem 3.2.25. ***Slutsky's Theorem***

Let $(S_{(1)}, S_{(2)}, S_{(3)}, ...)$ and $(T_{(1)}, T_{(2)}, T_{(3)}, ...)$ be sequences of random variables. Let T be a random variable and $c \in \mathbb{R}$. If $S_{(n)} \xrightarrow{p} c$ and $T_{(n)} \xrightarrow{d} T$, then:

- $S_{(n)} + T_{(n)} \xrightarrow{d} c + T$.
- $S_{(n)} T_{(n)} \xrightarrow{d} cT$.
- $\frac{T_{(n)}}{S_{(n)}} \xrightarrow{d} \frac{T}{c}$, provided that $c \neq 0$.

Definition 3.2.26. ***Asymptotic Unbiasedness***

An estimator $\hat{\theta}$ is asymptotically unbiased for θ if $\hat{\theta} \xrightarrow{d} T$, where $E[T] = \theta$.

Definition 3.2.28. ***Asymptotic Normality***

An estimator $\hat{\theta}$ is asymptotically normal if $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \phi^2)$, for some finite $\phi > 0$.

Definition 3.2.29. ***Asymptotic Standard Error***

For an estimator $\hat{\theta}$ such that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} T$, the asymptotic standard error of $\hat{\theta}$ is $V[T]$.

Definition 3.2.30. ***Asymptotic MSE***

For an estimator $\hat{\theta}$ such that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} T$, the asymptotic MSE of $\hat{\theta}$ is $E[T^2]$.

Definition 3.2.31. ***Asymptotic Relative Efficiency***

Let $\hat{\theta}_A$ and $\hat{\theta}_B$ be estimators of θ. Then $\hat{\theta}_A$ is asymptotically more efficient than $\hat{\theta}_B$ if it has a lower asymptotic MSE.

Definition 3.2.32. ***Consistency of Sampling Variance and Standard Error Estimators***

For an estimator $\hat{\theta}$ such that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} T$, a sampling variance estimator $\hat{V}[\hat{\theta}]$ is consistent if

$$n\hat{V}[\theta] \xrightarrow{p} V[T]$$

and a standard error estimator $\sqrt{\hat{V}[\hat{\theta}]}$ is consistent if

$$\sqrt{n}\sqrt{\hat{V}[\hat{\theta}]} \xrightarrow{p} V[T]$$

Theorem 3.2.33. ***Estimating the Sampling Variance of the Sample Mean***

For i.i.d. random variables $X_1, X_2, ..., X_n$ with finite variance $V[X] > 0$, let $\hat{V}[\bar{X}] = \frac{\hat{V}[X]}{n}$. Then:

- $E\left[\hat{V}[\bar{X}]\right] = V[\bar{X}]$.
- $n\hat{V}[\bar{X}] - nV[\bar{X}] = n\hat{V}[\bar{X}] - V[X] \xrightarrow{p} 0$.

Definition 3.2.34. ***Standard Error of the Sample Mean***

For i.i.d. random variables $X_1, X_2, ..., X_n$, the standard error of the sample mean is:

$$\sigma\left[\bar{X}\right] = \sqrt{V\left[\bar{X}\right]}$$

Theorem 3.2.35. ***Consistency of the Standard Error of the Sample Mean Estimator***

For i.i.d. random variables $X_1, X_2, ..., X_n$ with finite variance $V[X] > 0$,

$$\sqrt{n}\hat{\sigma}\left[\bar{X}\right] - \sqrt{V\left[X\right]} = \sqrt{n}\hat{\sigma}\left[\bar{X}\right] - \sqrt{n}\sigma\left[\bar{X}\right] \xrightarrow{p} 0$$

.