Random Variables

Statistical Methods in Political Research I

Yuki Shiraito

University of Michigan

Fall 2019

Random Variables

- Data are numbers
- How do we link sample spaces and events to numbers?

- Implicitly we have used:
  - Dice roll: Each outcome has a number
  - Falling stick: Azimuth degrees from 0 to 360
  - Survey responses: Yes as 1, No as 0
  - Supreme court: Number of judges voting for the plaintiff

- Random variable: A *random variable* X is a function
$$X : \Omega \to \mathbb{R}$$
such that for any real number $x \in \mathbb{R}$, $\{\omega \mid X(\omega) \leq x\} \in \mathcal{F}$

- Convention:
  1. Uppercase letter such as X stands for r.v.
  2. Lowercase letter such as x stands for a *realized value* of r.v.

Remarks on Random Variables

- Random variable is a *function*:
  - Takes an outcome in the sample space as an argument
  - Gives a single value assigned to each outcome
  - May give a common value for multiple outcomes

- Can consider different r.v.s for the same probability space:
  - Dice roll:
    - Numbers on the dice
    - -1 if 1 on the dice, 1 if 6 on the dice, 0 otherwise
    - 1 if an even number on the dice, 0 if an odd number on the dice
  - Survey responses:
    - 1 if "yes", 0 if "no" for each response
    - Number of respondents who answer "yes" (sum of the above)
    - Number of times a respondent answers "yes" (multiple responses)

- In applications, it is important to find a useful r.v.

Distribution

- *Distribution* of a random variable:
    - Let C be a subset of $\mathbb{R}$ such that $\{\omega \mid X(\omega) \in C\}$ is an event
        1. $\mathbb{P}(X \in C) \equiv \mathbb{P}(\{\omega \mid X(\omega) \in C\})$
        2. The *distribution* of X: The collection of $\mathbb{P}(X \in C)$ for all possible C
- Distribution of X can be considered as a probability measure:
    1. Sample space: $\mathbb{R}$
    2. Set of events: Set of all possible C
    3. Probability measure: $\mathbb{P}(X \in C)$
- Betting on even or odd numbers from a dice roll:
    - Sample space: 6 faces of a dice
    - Events: $\emptyset$, even, odd, all
    - Probability measures: $0, 1/2, 1/2, 1$
    - Random variable: $X(\omega) = 1$ if even, $X(\omega) = 0$ if odd
    - $\mathbb{P}(X \leq 0) \equiv \mathbb{P}(\text{odd}) = 1/2, \mathbb{P}(X > 0) \equiv \mathbb{P}(\text{even}) = 1/2$
    - $C \in \{\emptyset, \{r \in \mathbb{R} \mid r \leq 0\}, \{r \in \mathbb{R} \mid r > 0\}, \mathbb{R}\}$

- Will directly work with r.v.: Write X instead of $X(\omega)$
- Probability space is hidden behind r.v., but it's there

Cumulative Distribution Function

- Generally, there are a huge number of C
- Need a simple way to describe a distribution

- Cumulative distribution function: The *cumulative distribution function (c.d.f.)* of a r.v. X, denoted by F_X, is a function $F_X : \mathbb{R} \to [0, 1]$ such that
$$F_X(x) = \mathbb{P}(X \leq x)$$
- Remember the definition of r.v.: $\{\omega \mid X(\omega) \leq x\} \in \mathcal{F}$ for any $x \in \mathbb{R}$

- Example of c.d.f.:
  - Betting on even or odd numbers on a dice roll
  - Dice roll in the Nigeria Survey
  - Stick fall in the Nigeria Survey

- Uniqueness of c.d.f.: Let X have c.d.f. F and Y have c.d.f. G. If $F(x) = G(x)$ for all $x \in \mathbb{R}$, then $\mathbb{P}(X \in C) = \mathbb{P}(Y \in C)$ for all possible C.

Valid C.D.F.

- Properties of c.d.f.: A function $F : \mathbb{R} \to [0, 1]$ is a c.d.f. for some probability $\mathbb{P}$ if and only if F satisfies the following three conditions:
    1. F is non-decreasing: $x_1 < x_2$ implies that $F(x_1) \leq F(x_2)$
    2. F is normalized:
    $$\lim_{x \to -\infty} F(x) = 0$$
    $$\lim_{x \to \infty} F(x) = 1$$
    3. F is right-continuous: $F(x) = \lim_{y \downarrow x} F(y)$ for all x

  Proof. Jerry's section and future problem set.

- *Any* function satisfying the three conditions can be a c.d.f.
- Not necessarily well known

Discrete Random Variable

- Discrete r.v.: X is *discrete* if it takes only countably many values
- Probability function: For a discrete r.v. X, the *probability (mass) function (p.f.)* of X, denoted by $f_X : \mathbb{R} \to [0, 1]$, is defined by $f_X(x) \equiv \mathbb{P}(X = x)$
- Support of X: $\{x \in \mathbb{R} \mid f_X(x) > 0\}$
- C.d.f. and p.f.:

$$F_X(x) = \sum_{\{y \mid f_X(y) > 0 \,\wedge\, y \leq x\}} f_X(y)$$

$$f_X(x) = \lim_{y \downarrow x} F_X(y) - \lim_{y \uparrow x} F_X(y)$$

- X is discrete $\Leftrightarrow$ c.d.f. of X is a step function
- Valid p.f.: The p.f. of X with its support $\{x_1, \ldots\}$ must satisfy the following two conditions:
    1. f is non-negative: $f_X(x) \geq 0$
    2. f sums to 1: $\sum_{i=1}^{\infty} f_X(x_i) = 1$

Bernoulli Distribution

- *Bernoulli trial*: Random realization of a "success" or a "failure"
  - Survey response to a yes/no question
  - Yea/nay vote by a legislator, judge, representative, . . .
  - Any binary feature (e.g. democracy, below/above a threshold, etc.)
- X follows a Bernoulli distribution with support $\{0, 1\}$:

$$F_X(x) = \begin{cases} 0 & (x < 0) \\ 1 - p & (0 \leq x < 1) \\ 1 & (1 \leq x) \end{cases}$$

$$f_X(x) = p^{1\{x=1\}}(1 - p)^{1\{x=0\}}$$

  - *Parameter* of the Bernoulli distribution: $p = \mathbb{P}(X = 1)$
  - $1\{\cdot\}$: Indicator function
  - Denoted by: $X \sim \text{Bern}(p)$
- C.d.f. and p.f. of known distributions:
  - Values of X and parameters
  - $F_X(x; \theta)$, $f_X(x; \theta)$

Binomial Distribution

- Sum of "successes" in n independent Bernoulli trials
  - Nigeria survey: Number of "yes" answers if asked multiple times
  - Conflict example in Pset 2: Number of battles a country wins
  - Shop owner: Number of customers who give 3+ stars on Yelp
- X follows a Binomial distribution with support $\{0, 1, \dots\}$:

$$F_X(x; n, p) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$$

$$f_X(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

  - *Parameters* of the Binomial distribution: p and n
  - $\lfloor x \rfloor$: the greatest integer less than or equal to x
  - Denoted by: $X \sim \text{Binom}(n, p)$

- Story gives the p.f. of the Binomial distribution
- Binomial r.v. is the sum of Bernoulli r.v.
- Will revisit the transformation of r.v.s

Continuous Random Variable

- Probability density function: If there exists a function $f_X : \mathbb{R} \to \mathbb{R}^+$ such that for any $a, b \in \mathbb{R}$ with $a \leq b$,
$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx$$
and $\int_{-\infty}^{\infty} f_X(x)dx = 1$, then $f_X(x)$ is called the *probability density function (p.d.f.)* of X
- A random variable X is continuous if there exists a p.d.f. of X
- Support of X: $\{x \in \mathbb{R} \mid f_X(x) > 0\}$
- C.d.f. and p.d.f.:
$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$
$$f_X(x) = F'_X(x) \text{ for any } x \text{ at which } F_X \text{ is differentiable}$$
- Types of r.v.:
  1. Discrete: Support is countable
  2. Continuous: P.d.f. exists $\Rightarrow$ support is uncountable
  3. Neither: Support is uncountable but p.d.f. does not exist

Uniform Distribution

- "Completely random" number over a continuous interval
  - Nigeria survey: Direction a stick falls
- X follows the Uniform distribution on the interval $[a, b]$:

$$F_X(x; a, b) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b < x \end{cases}$$

$$f_X(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

  - Parameters of the Uniform distribution: a and b
  - Denoted by: $X \sim \text{Unif}(a, b)$
  - Commonly on the unit interval $[0, 1]$

- Not many real-world examples, unless artificially created
- Useful tool for modeling and simulations

Functions of Random Variables

- How to generate random numbers?
  - Online survey: Randomly switch questions
  - Experiment: Randomly assign treatment or control
- Function of r.v. is also r.v.:
  - If $X : \Omega \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, then $g \circ X : \Omega \to \mathbb{R}$
  - $Y(\omega) \equiv g(X(\omega))$ is r.v.
- If X is discrete:
  - P.f. of Y: $f_Y(y) \equiv \mathbb{P}(Y = y) = \mathbb{P}(g(X) = y) = \sum_{\{x \mid g(x) = y\}} f_X(x)$
  - E.g., $Y = n - X$ where $X \sim \text{Binom}(n, p)$: $f_Y(y) = \binom{n}{n-y} p^{n-y}(1-p)^y$
- If X is continuous:
  - C.d.f. of Y:

$$F_Y(y) \equiv \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \in \{x \mid g(x) \leq y\})$$
$$= \int_{\{x \mid g(x) \leq y\}} f_X(x) dx$$

  - E.g., $Y = X^2$ where $X \sim \text{Unif}(0, 1)$: $F_Y(y) = \int_0^{\sqrt{y}} 1 dx$

Inverse-CDF Method

- Increasing function of r.v.:
  - $g : \mathbb{R} \to \mathbb{R}$ is increasing: $a < b$ implies that $g(a) < g(b)$
  - If g is increasing, then $F_Y(y) = F_X(g^{-1}(y))$
- Quantile function: The *quantile function (q.f.)* (a.k.a. inverse c.d.f.) of X, denoted by $Q_X : [0, 1] \to \mathbb{R}$, is a function
$$Q_X(u) \equiv \inf\{x \mid F_X(x) > u\}$$
- Inverse-CDF method: Let X is an r.v. with $Q_X(\cdot)$, $U \sim \text{Unif}(0, 1)$, and $Y = Q_X(U)$. Then, $F_Y(y) = F_X(y)$
  - $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(Q_X(U) \leq y) = \mathbb{P}(U \leq F_X(y)) = F_X(y)$
  - $\inf\{x \mid F_X(x) > U\} \leq y \Leftrightarrow U \leq F_X(y)$: **Proof.** Jerry's section
- Generating random numbers whose c.d.f. is F_X:
  - Generate $U \sim \text{Unif}(0, 1)$
  - Transform by $X = Q_X(U)$
- Special Case:
  - F_X is increasing $\Rightarrow Q_X(U) = F_X^{-1}(U)$
  - $F_U(Q_X^{-1}(x)) = F_U((F_X^{-1})^{-1}(x)) = F_U(F_X(x)) = F_X(x)$

Multivariate Random Variables

- Multiple r.v.s:
  - Dice roll: Indicator of each number on the dice
  - Survey: Responses by many respondents

- Joint c.d.f.: The *joint c.d.f.* of a random vector $\mathrm{X} \equiv (X_1, \dots, X_n)$, denoted by $F_{\mathrm{X}} : \mathbb{R}^n \to [0, 1]$, is a function
$$F_{\mathrm{X}}(\mathrm{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$
where $\mathrm{x} \equiv (x_1, \dots, x_n)$
- Dice roll and indicator: $X_i \equiv 1\{i \text{ shows on the dice}\}$
  - X_i is either 0 or 1
  - $F_{\mathrm{X}}(\mathrm{x}) = j/6$ where j is the number of 1 in x

- Joint p.f.: Let $X_1, \dots, X_n$ be discrete r.v.s. The *joint p.f.* of X, denoted by $f_{\mathrm{X}} : \mathbb{R}^n \to [0, 1]$ is a function
$$f_{\mathrm{X}}(\mathrm{x}) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$
- Dice roll and indicator, again:
  - $f_{\mathrm{X}}(\mathrm{x}) = 1/6$ for any x such that only one element is 1
  - $f_{\mathrm{X}}(\mathrm{x}) = 0$ otherwise

Multinomial Distribution

- The number of times each "category" appears in n trials
  - Multiple dice rolls: How many times each number shows
  - Responses to multiple choice/count questions
  - Word counts in a document (Problem Set 3)

- X follows a Multinomial distribution:
  - Joint p.f.: For non-negative integers $x_1, \ldots x_K$,

$$f_{\mathrm{X}}(\mathrm{x}; n, \mathrm{p}) = \begin{cases} \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k} & \text{when } \sum_{i=1}^{k} x_i = n \\ 0 & \text{otherwise} \end{cases}$$

  - Parameters of the Multinomial distribution: n and p
  - $p_1 + \cdots + p_k = 1$
  - Denoted by: $\mathrm{X} \sim \mathrm{Multi}(n, \mathrm{p})$

- Multinomial is Binomial if $k = 2$
- Multinomial is the sum of indicators for each trial

Multivariate Uniform

- Joint p.d.f.: For a random vector $\mathrm{X} = (X_1, \ldots, X_n)$, if there exists a function $f_{\mathrm{X}} : \mathbb{R}^n \to \mathbb{R}^+$ such that for any set $C \subset \mathbb{R}^n$,
$$\mathbb{P}(\mathrm{X} \in C) = \int \underset{C}{\cdots} \int f_{\mathrm{X}}(\mathrm{x}) dx_1 \ldots dx_n$$
and $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathrm{X}}(\mathrm{x}) dx_1 \ldots dx_n = 1$, then $f_{\mathrm{X}}(\mathrm{x})$ is called the *joint p.d.f.* of X

- $\mathrm{X} = (X_1, X_2)$ follows a Uniform distribution over $[0,1] \times [0,1]$:
  - Joint p.d.f.:
$$f_{\mathrm{X}}(\mathrm{x}) = \begin{cases} 1 & 0 \le x_1 \le 1,\ 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$
  - Joint c.d.f.:
$$F_{\mathrm{X}}(\mathrm{x}) = \begin{cases} 0 & x_1 < 0 \text{ or } x_2 < 0 \\ x_1 x_2 & 0 \le x_1 \le 1 \text{ and } 0 \le x_2 \le 1 \\ 1 & x_1 > 1 \text{ and } x_2 > 1 \end{cases}$$

Marginal Distribution

- Marginal c.d.f.: Let $\mathrm{X} = (X_1, \ldots, X_n)$ be a random vector and F_{X} be its joint c.d.f. Then,
  $F_{X_i}(x_i) = \lim_{x_1 \to \infty} \ldots \lim_{x_{i-1} \to \infty} \lim_{x_{i+1} \to \infty} \ldots \lim_{x_n \to \infty} F_{\mathrm{X}}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots x_n)$
  F_{X_i} is called the *marginal c.d.f.* of X_i: **Proof**. $n = 2$ case
- F_{X_i} is a valid c.d.f.: The *marginal distribution* of X_i

- Marginal p.f.: If the marginal distribution of X_i is discrete, the *marginal p.f.* of X_i is defined by
  $$f_{X_i}(x) \equiv \mathbb{P}(X_i = x_i) = F_{X_i}(x_i) - \lim_{y \uparrow x_i} F_{X_i}(y)$$
- If a joint p.f. $f_{\mathrm{X}}(\mathrm{x})$ exists, $f_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} f_{\mathrm{X}}(\mathrm{x})$

- Marginal p.d.f.: If the marginal c.d.f. F_{X_i} has a p.d.f. f_{X_i}, it is called a *marginal p.d.f.* of X_i
- If a joint p.d.f. $f_{\mathrm{X}}(\mathrm{x})$ exists,
  $f_{X_i}(x_i) = \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_n} f_{\mathrm{X}}(\mathrm{x}) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$

Independence

- Independent r.v.s: **Very** important in data analysis
- Independence of r.v.s: R.v.s $X_1, \dots, X_n$ are *independent* if and only if for any subsets $C_1, \dots C_n$ of $\mathbb{R}$,
$$\mathbb{P}(X_1 \in C_1, \dots, X_n \in C_n) = \mathbb{P}(X_1 \in C_1) \dots \mathbb{P}(X_n \in C_n)$$
- Notation: $X_i \perp\!\!\!\perp X_j$
- Knowing the value of X_i does not help predict X_j

- Connection with marginal distribution:
  - $X_1, \dots, X_n$ are independent if and only if for any $x_1, \dots, x_n \in \mathbb{R}$
$$F_{\mathrm{X}}(\mathrm{x}) = \prod_{i=1}^{n} F_{X_i}(x_i)$$
  - If a joint p.f. $f_{\mathrm{X}}(\mathrm{x})$ exists, $f_{\mathrm{X}}(\mathrm{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$
  - If a joint p.d.f. $f_{\mathrm{X}}(\mathrm{x})$ exists, $f_{\mathrm{X}}(x) = \prod_{i=1}^{n} f_{X_i}(x_i)$

I.i.d. and Random Sample

- Benchmark model of data generating process
  1. Data points are independent random variables
  2. All data points follow a common distribution
- Independent and identically distributed: $X_1, \ldots, X_n$ are *i.i.d. (independent and identically distributed)* if and only if they are independent and each has the same marginal distribution with c.d.f. F
- Notation: $X_i \overset{\text{i.i.d.}}{\sim} F$ or $X_i \overset{\text{i.i.d.}}{\sim} f$
- $(X_1, \ldots, X_n)$ is called *a random sample of size n from F*

- Random sampling for opinion poll:
  - Population N, Dem supporters $m_D < N$, $p \equiv m_D/N$
  - X_i: 1 if person i is Dem supporter, 0 otherwise
  - i is randomly chosen:
    1. X_i is i.i.d. Bernoulli with p with the *super population* assumption
    2. X_i is independent but not indentically distributed under the *finite population* assumption

Convolution

- Sum of two random variables is called *convolution*
- Convolution: Let X_1 and X_2 be independent random variables and $Y \equiv X_1 + X_2$. Then, the distribution of Y is called the *convolution* of the distributions of X_1 and X_2.
- If both X_1 and X_2 are discrete,

$$f_Y(y) = \sum_{x_1=-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y - x_1)$$

- If both X_1 and X_2 are continuous,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1$$

- Convolution of Bernoulli r.v.s with common p is Binom$(2, p)$:

$$f_Y(y) = \sum_{x_1=0}^{1} f_{X_1}(x_1) f_{X_2}(y - x_1) = \begin{cases} (1-p)^2 & (y = 0) \\ 2p(1-p) & (y = 1) \\ p^2 & (y = 2) \end{cases}$$

Conditional Distribution

- In many studies, *prediction* is of interest
    - Vote choice given ethnicity, gender, age, etc...
    - Attitude toward immigration given occupation
    - Economic growth/conflict behavior given regime type
- *Regression* (covered in 699) is the most important method
- *Conditional distribution* is the idea behind regression
- Conditional c.d.f.: Let $\mathrm{X} \equiv (X_1, \ldots, X_n)$ have the joint distribution with c.d.f. F_{X}. Then,

$$
\begin{aligned}
&F_{\mathrm{X}_{1:i}|\mathrm{X}_{(i+1):n}}(x_1, \ldots, x_i \mid \mathrm{X}_{(i+1):n} \in \times_{j=i+1}^{n} C_j) \\
\equiv &\mathbb{P}(X_1 \leq x_1, \ldots, X_i \leq x_i \mid X_{i+1} \in C_{i+1}, \ldots, X_n \in C_n) \\
= &\frac{F_{\mathrm{X}}(x_1, \ldots, x_n)}{\mathbb{P}(X_{i+1} \in C_{i+1}, \ldots, X_n \in C_n)}
\end{aligned}
$$

  for $C_j \subset \mathbb{R}, j = i + 1, \ldots, n$, is called the *conditional c.d.f. of* $\mathrm{X}_{1:i}$ *given that* $X_{i+1} \in C_{i+1}, \ldots, X_n \in C_n$
- Conditional c.d.f. uniquely defines the *conditional distribution of* $\mathrm{X}_{1:i}$ *given that* $\mathrm{X}_{(i+1):n} \in \times_{j=i+1}^{n} C_j$

Conditional P.(d.)f. and Hybrid Random Vectors

- If X is discrete, then there exists the *conditional p.f.*:
$$f_{\mathrm{X}_{1:i}|\mathrm{X}_{(i+1):n}}(x_1,\ldots,x_i \mid x_{i+1},\ldots,x_n) = \frac{f_{\mathrm{X}}(x_1,\ldots,x_n)}{f_{\mathrm{X}_{(i+1):n}}(x_{i+1},\ldots,x_n)}$$
- If X is continuous, then there exists the *conditional p.d.f.*:
$$f_{\mathrm{X}_{1:i}|\mathrm{X}_{(i+1):n}}(x_1,\ldots,x_i \mid x_{i+1},\ldots,x_n) = \frac{f_{\mathrm{X}}(x_1,\ldots,x_n)}{f_{\mathrm{X}_{(i+1):n}}(x_{i+1},\ldots,x_n)}$$
- Independence $\Leftrightarrow f_{X_1|X_2}(x_1 \mid x_2) = f_{X_1}(x_1)$
- Joint p.(d.)f. = cond. p.(d.)f. $\times$ marg. p.(d.)f.
- Joint p.f.-p.d.f. of a hybrid random vector: Let $\mathrm{X}_{(i+1):n}$ have a marginal p.d.f. (p.f.) $f_{\mathrm{X}_{(i+1):n}}$ and $\mathrm{X}_{1:i}$ have the conditional p.f. (p.d.f.) $f_{\mathrm{X}_{1:i}|\mathrm{X}_{(i+1):n}}$. Then, we define the *joint p.f.-p.d.f.* of X as
$$\begin{aligned}&f_{\mathrm{X}}(x_1,\ldots,x_n)\\ \equiv &f_{\mathrm{X}_{1:i}|\mathrm{X}_{(i+1):n}}(x_1,\ldots,x_i \mid x_{i+1},\ldots,x_n) f_{\mathrm{X}_{(i+1):n}}(x_{i+1},\ldots,x_n)\end{aligned}$$

Uniform-Binomial Model

- A popular model in Bayesian statistics:
  1. Proportion of Dem supporters drawn from the Uniform
  2. Random sample of size n for a survey on partisanship
- Data generating process:
  1. Proportion of Dem supporters: $X_1 \sim U[0,1]$
  2. Number of Dem supporters in sample: $X_2 \sim \text{Binom}(n, X_1)$
- Joint p.f.-p.d.f.:

$$f_{X_1,X_2}(x_1,x_2) = f_{X_2|X_1}(x_2 \mid x_1) f_{X_1}(x_1) = 1 \times \binom{n}{x_2} x_1^{x_2}(1-x_1)^{n-x_2}$$

- Marginal p.d.f. and p.f.:
  1. Marginal p.d.f. of X_1:

$$f_{X_1}(x_1) = \begin{cases} 1 & (0 \leq x_1 \leq 1) \\ 0 & (\text{otherwise}) \end{cases}$$

  2. Marginal p.f. of X_2: Letting $B(\cdot,\cdot)$ be the Beta function

$$f_{X_2}(x_2) = \binom{n}{x_2} \int_0^1 x_1^{x_2}(1-x_1)^{n-x_2} dx_1 = \binom{n}{x_2} B(x_2+1, n-x_2+1)$$

Bayes' Theorem for Random Variables

- The number of Dem supporters in sample is known
- Need to estimate the proportion of Dems in population

- Bayes' theorem for r.v.s: Let (X_1, X_2) has a joint p.f., p.d.f., or p.f.-p.d.f. Then,

$$f_{X_1}(x_1) = \frac{f_{X_2|X_1}(x_2 \mid x_1) f_{X_1}(x_1)}{f_{X_2}(x_2)}$$

- Law of total probability for r.v.s: Let (X_1, X_2) has a joint p.f., p.d.f., or p.f.-p.d.f. Then,

$$f_{X_2}(x_2) = \sum_{x_1} f_{X_2|X_1}(x_2 \mid x_1) f_{X_1}(x_1) \quad (X_1 \text{ is discrete})$$

$$f_{X_2}(x_2) = \int_{x_1} f_{X_2|X_1}(x_2 \mid x_1) f_{X_1}(x_1) dx_1 \quad (X_1 \text{ is continuous})$$

- *Posterior* p.d.f. of X_1 given X_2 in the Uniform-Binomial:

$$f_{X_1|X_2}(x_1 \mid x_2) = \frac{x_1^{x_2}(1 - x_2)^{n - x_2}}{B(x_2 + 1, n - x_2 + 1)}$$

Posterior P.d.f.