Random Variables

Ye Wang
University of North Carolina at Chapel Hill

Mathematics and Statistics For Political Research
POLI783

Random variables

- We can always study the probability of random events.
- But a more convenient approach is to assign real numbers to these events.
- For instance, instead of analyzing outcomes in $\{H, T\}$, we can define a function:

$$Y = f(X) = \begin{cases} 1 & \text{if } X = H, \\ 0 & \text{if } X = T. \end{cases}$$

- Now, the two outcomes are represented by two numbers.
- This transformation allows us to analyze similar events using the same function.
- E.g., we can apply the same coding to electoral outcomes:

$$Y = f(X) = \begin{cases} 1 & \text{if } X = \text{Win}, \\ 0 & \text{if } X = \text{Loss}. \end{cases}$$

- The same function can describe different types of events across domains.

Random variables

- These functions from the σ-algebra ($\mathcal{F}$) to the real line ($\mathbb{R}$) are called random variables.
- They help us understand experiments that can have infinite outcomes.
- As social scientists, we do this all the time!
- We often "code" various outcomes as numeric variables in our research.
- E.g., Republicans may be coded as 1 and Democrats as 0.
- This coding transforms qualitative events into numeric objects we can model.
- Probabilities originally assigned to events are now transferred to the values of the random variable.
- That's why they are called "random" variables.

Random variables

- Different random variables can be defined for the same experiment.
- If we randomly select two senators, there will be four outcomes:

$$S = \{DD, DR, RD, RR\}.$$

- We can define X as the number of Democrats.
- $X \in \{0, 1, 2\}$.
- Or, we can define Y as

$$Y = \begin{cases} 1 \text{ if } DD, \\ 0 \text{ otherwise.} \end{cases}$$

Random variables

- There are two main types of random variables: discrete and continuous.
- Discrete random variables describe sample spaces where the number of outcomes is finite or countably infinite.
- A random variable X is discrete if the values it takes with positive probability are finite:

$$X \in \{x_1, x_2, \dots, x_K\}$$

  or countably infinite:

$$X \in \{x_1, x_2, \dots\}.$$

- The collection of x where $\mathbb{P}(X = x) > 0$ is called the support of X.
- We denote it as $\mathcal{X}$.

Discrete random variables

- We can describe a discrete random variables using its probability mass function (p.m.f.):

$$p_X(x) = \mathbb{P}(X = x).$$

- $p_X(x)$ is well-defined for any x since $\{X = x\}$ is a random event.
- E.g., $\{X = 2\}$ means there are two Democrats in our sample of senators.
- $p_X(x) > 0$ for any x in X's support and $p_X(x) = 0$ otherwise.
- $\mathbb{P}(X \in S) = \sum_{x \in S} p_X(x)$ and $\sum_{x \in \mathcal{X}} p_X(x) = 1$.
- These properties come from the axioms of probability.

Bernoulli distribution

- A random variable X has a Bernoulli distribution with parameter p if

$$p_X(1) = \mathbb{P}(X = 1) = p, p_X(0) = \mathbb{P}(X = 0) = 1 - p.$$

- It describes the success or failure of a single experiment.
- Flipping a coin, launching a policy, declaring a war, etc.
- We can describe the occurrence of any random event A using the Bernoulli distribution:

$$\mathbf{1}\{A\} \sim Bern(p), \text{ with } p = \mathbb{P}(A).$$

- With different ps, we have different distributions.

Treatment assignment

- Bernoulli distributions are commonly used in randomized trials.
- Consider a randomized trial with three participants.
- We flip a coin to decide who should receive the treatment (e.g., a new medicine).
- The participant is treated with the medicine if it is a head.
- We often use D to denote the treatment status:

$$D = \begin{cases} 1, & p, \\ 0, & 1 - p. \end{cases}$$

Treatment assignment

- Suppose $p = \frac{1}{2}$.
- We use X to represent the number of treated units.
- $X \in \{0, 1, 2, 3\}$.
- What is the p.m.f. for X?

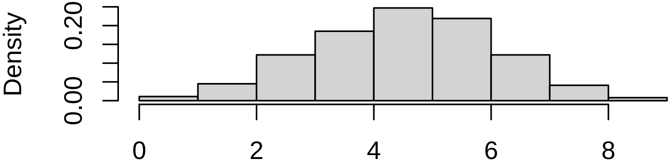
$$X = \begin{cases} 0, & \text{if } \{T, T, T\}, p = \frac{1}{8}, \\ 1, & \text{if } \{T, T, H\}, \{H, T, T\}, \{T, H, T\}, p = \frac{3}{8}, \\ 2, & \text{if } \{T, H, H\}, \{H, T, H\}, \{H, H, T\}, p = \frac{3}{8} \\ 3, & \text{if } \{H, H, H\}, p = \frac{1}{8}. \end{cases}$$

Binomial distribution

- In this example, X represents the number of successes in 3 Bernoulli trials.
- With n Bernoulli trials in total, X obeys the binomial distribution: $X \sim Bin(n, p)$.
- The probability for k trials to succeed is

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

- This gives the p.m.f of X.



Poisson distribution

- A discrete r.v. can take infinitely countable values.
- A r.v. X follows the Poisson distribution if

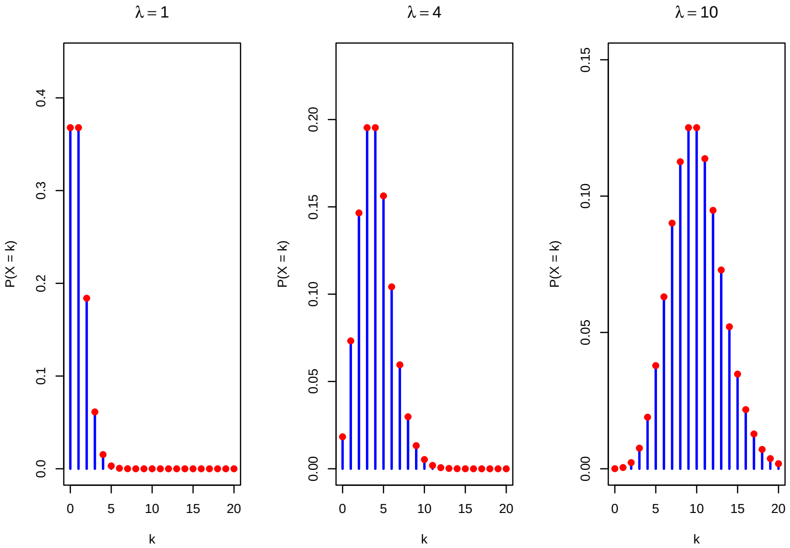
$$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

- We can see that

$$\sum_{x \in \mathcal{X}} p_X(x) = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} * e^{\lambda} = 1.$$

- If X represents the number of events (emails, protests, wars) in a fixed time interval (a day, a year, a decade), then X often follows the Poisson distribution.
- It can be seen as the limit of the binomial distribution as $n \to \infty$ and $p \to 0$, with $\lambda = \lim(np)$.

Poisson distribution



Cumulative distribution functions

- The cumulative distribution function (c.d.f.) of a random variable X is defined as

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

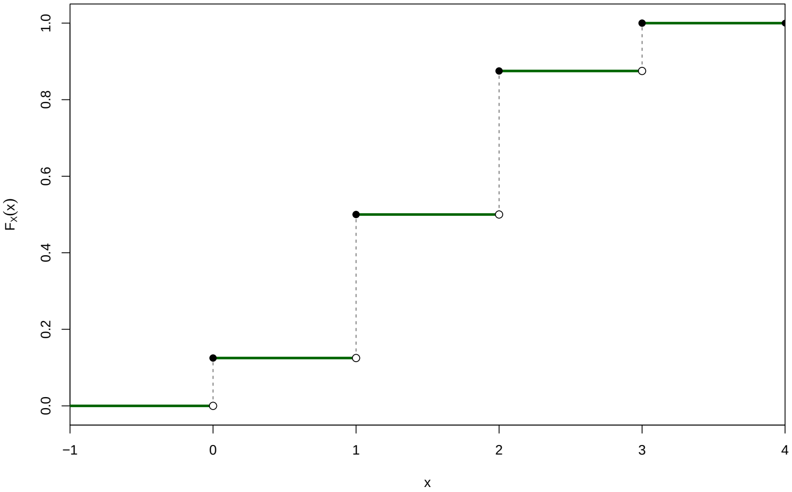
- For discrete random variables, we know that

$$F_X(x) = \sum_{x' \leq x} \mathbb{P}(X = x').$$

- Consider X in the previous example:

$$F_X(x) = \begin{cases} 0, \text{ if } X < 0, \\ \frac{1}{8}, \text{ if } 0 \leq X < 1, \\ \frac{1}{2}, \text{ if } 1 \leq X < 2, \\ \frac{7}{8}, \text{ if } 2 \leq X < 3, \\ 1, \text{ if } X \geq 3. \end{cases}$$

Cumulative distribution functions



Properties of the c.d.f.

- The c.d.f. allows us to find the probability of any region:

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$$
$$\mathbb{P}(X > a) = 1 - F_X(a).$$

- Obviously, $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
- In addition, $F_X(x)$ is non-decreasing and right-continuous:

$$F_X(x) \leq F_X(x') \text{ for any } x \leq x',$$
$$F_X(x) = \lim_{x' \to x+} F_X(x').$$

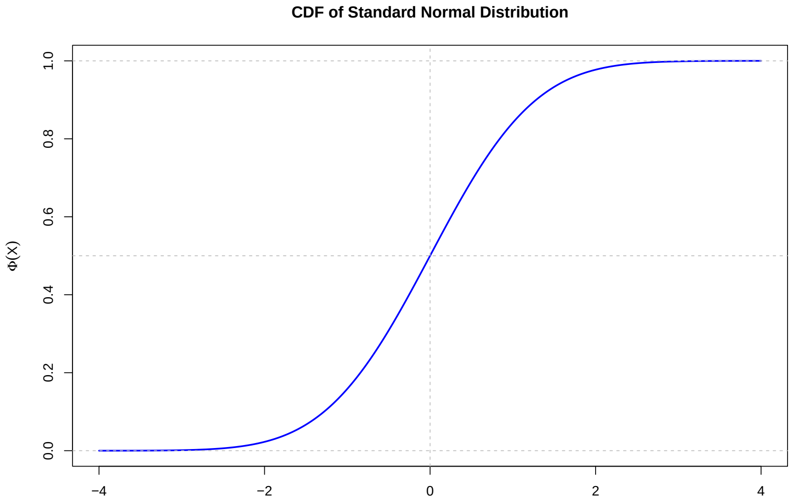
- These properties hold due to the definition of $F_X(x)$.

Continuous random variables

- Continuous random variables can take any real number (in an interval) as their values.
- They cannot be described by the p.m.f.
- Otherwise, $\sum_{x\in[0,1]} p_X(x) = \infty$.
- It violates the axioms of probability.

Continuous random variables

- Yet we can still describe a continuous r.v. using its c.d.f.



Continuous random variables

- A random variable X is continuous if its c.d.f. is a continuous function and there exists a function $f_X(x)$ such that:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

- $f_X(x)$ is known as the probability density function (p.d.f.) of X.
- By its definition, we know that $F_X'(x) = f_X(x)$ and

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$$

- Remember that $F_X(x)$ is non-decreasing, thus

$$f_X(x) \geq 0 \text{ for any } x,$$
$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Continuous random variables

- Using the definition of continuity, we know that

$$\mathbb{P}(X = x) = \lim_{\epsilon \to 0+} [F_X(x + \epsilon) - F_X(x)] = 0.$$

- The probability for X to take any value x is 0.
- $\mathbb{P}(X = x)$ is $f(x)$'s definite integral at one point.
- It makes more sense to discuss the p.d.f. or the probability of a region for continuous random variables.

Uniform distribution

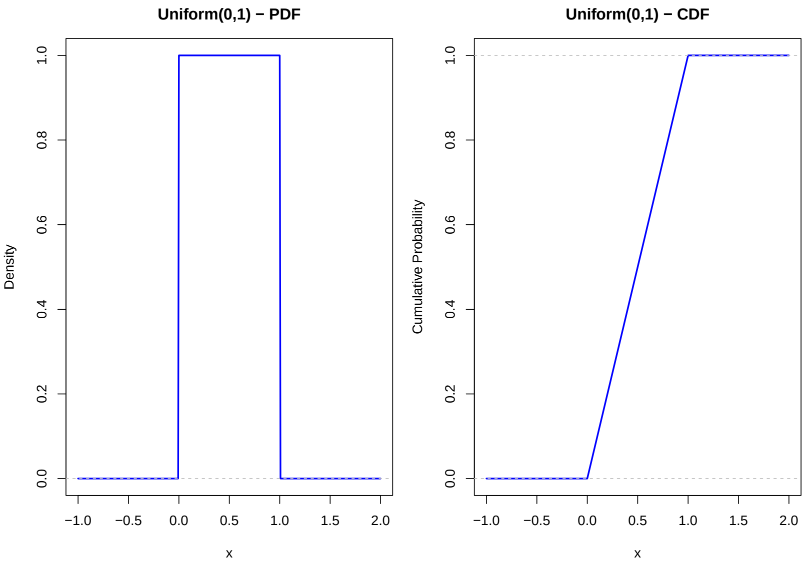
- If U's p.d.f. takes the same value on its support $[a, b]$:

$$f_U(u) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq u \leq b, \\ 0 & \text{otherwise,} \end{cases}$$

we say U obeys the uniform distribution: $U \sim Unif(a, b)$.

- We can see that for $a \leq c \leq d \leq b$, $P(c \leq X \leq d) = \frac{d-c}{b-a}$.
- A discrete random variable may also obey the uniform distribution.
- We can show that $cU + d \sim Unif(ca + d, cb + d)$.
- This is not true in general!
- If we know nothing about U but its support, we often assume that U is uniformly distributed (a flat prior).
- E.g., the proportion of dissidents in Cuba can be any number on $[0, 1]$.

Uniform distribution



Standard normal distribution

- If a r.v. Z's p.d.f. takes the following form:

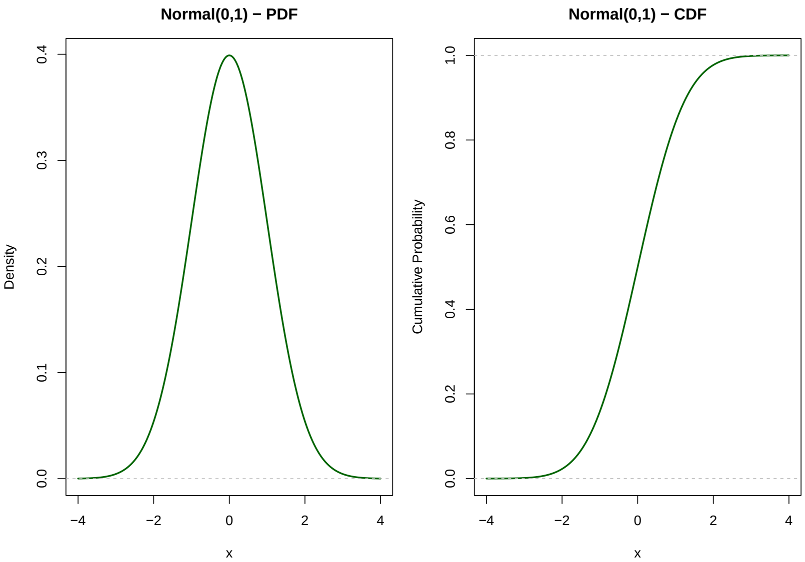
$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}},$$

  we say Z obeys the standard normal distribution: $Z \sim \mathcal{N}(0, 1)$.
- Note that Z's support is the real axis and $\phi_Z(z)$ is symmetric: $\phi_Z(z) = \phi_Z(-z)$.
- We can verify that $\int_{-\infty}^{\infty} \phi(z) dz = 1$.
- The c.d.f. of the standard normal distribution is usually denoted as

$$\Phi(z) = \int_{-\infty}^{z} \phi(t) dt.$$

- It is named as "normal" because the average of random noises often converges to this distribution.
- The foundation of statistical inference.

Standard normal distribution



Standard normal distribution

- If $Z \sim \mathcal{N}(0, 1)$, then
  - Roughly 68% of the distribution of Z is between -1 and 1
  - Roughly 95% of the distribution of Z is between -2 and 2
  - Roughly 99% of the distribution of Z is between -3 and 3
  - We often use these "critical values" for statistical inference

Independence of random variables

- We say two r.v.s X and Y are independent if

$$\mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\mathbb{P}(Y \le y)$$

- Similarly, this can be generalized to K r.v.s:

$$\mathbb{P}(X_1 \le x_1, \ldots, X_K \le x_K) = \prod_{k=1}^{K} \mathbb{P}(X_k \le x_k).$$

- Again, joint independence implies pairwise independence but not vice versa.
- For discrete r.v.s, this becomes

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

- We say a sequence of random variables, $\{X_1, X_2, \ldots, X_K\}$, are independent and identically distributed (i.i.d.) if they are jointly independent and obey the same distribution.

Functions of random variables

- For any r.v. X, we can apply a function to it and obtain

$$Y = g(X).$$

- The range of $g(\cdot)$ is the support of Y, and

$$\begin{aligned} F_Y(y) =& \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) \\ =& \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)), \end{aligned}$$

  if $g(\cdot)$ is strictly increasing.
- In general, $f_Y(y) \neq g(f_X(x))$, and we need to derive $F_Y(y)$ across regions where $g(\cdot)$ has different monotonicity.

Functions of random variables

- Consider the r.v. X we saw before:

$$X = \begin{cases} 0, & \text{if } \{T,T,T\}, p = \frac{1}{8}, \\ 1, & \text{if } \{T,T,H\},\{H,T,T\},\{T,H,T\}, p = \frac{3}{8}, \\ 2, & \text{if } \{T,H,H\},\{H,T,H\},\{H,H,T\}, p = \frac{3}{8} \\ 3, & \text{if } \{H,H,H\}, p = \frac{1}{8}. \end{cases}$$

- What if the p.m.f. for $Z = \mathbf{1}\{X > 2\}$?

$$Z = \begin{cases} 1, & \text{if } X \in \{3\}, p = \frac{1}{8}, \\ 0, & \text{if } X \in \{0,1,2\}, p = \frac{7}{8}. \end{cases}$$

Functions of random variables

- We can even define functions of multiple r.v.s:

$$Y = g(X_1, X_2, \ldots, X_n).$$

- E.g., if $X_i \sim Bern(p)$ and $X_i \perp\!\!\!\perp X_j$ for any i and j, then $Y = \sum_{i=1}^{n} X_i \sim Bin(n, p)$.
- If $Y = k$, then k variables in $\{X_i\}_{i=1}^{n}$ take the value of 1 and the rest $n - k$ ones take the value of 0, thus

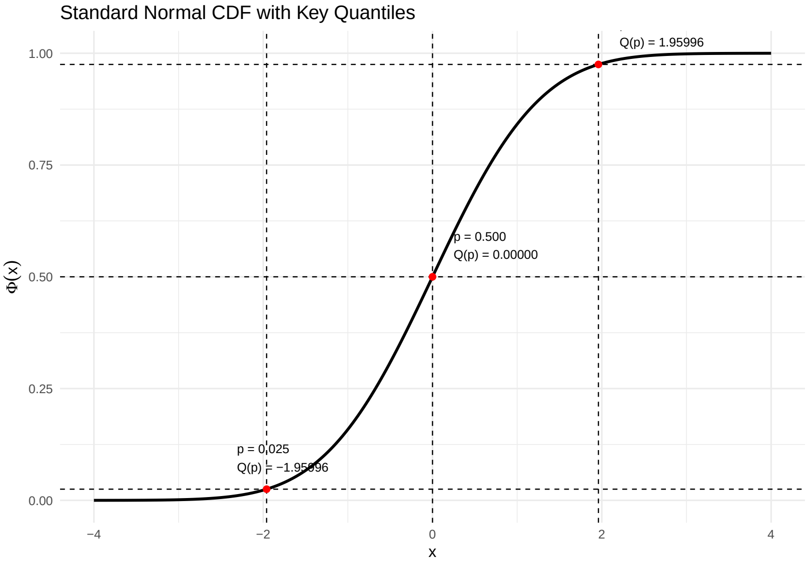
$$\mathbb{P}(Y = k) = \mathbb{P}\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- We can similarly prove that if $Y_1 \sim Bin(n, p)$ and $Y_2 \sim Bin(m, p)$, then $Y_1 + Y_2 \sim Bin(n + m, p)$.

Quantile function

- We call the inverse of the c.d.f. the quantile function and denote it as $F^{-1}(\alpha)$.
- If $F_X(x) = \mathbb{P}(X \leq x) = \alpha$, then $F^{-1}(\alpha) = x$.
- It maps a probability to a value of X.
- $F^{-1}(0.5)$ is known as the median.
- Quantiles are widely used in studies on income inequality.
- E.g., the income of the top 1% v.s. the income of the bottom 1%.
- We also use them to construct confidence intervals: $\Phi^{-1}(0.975) = 1.96$ and $\Phi^{-1}(0.025) = -1.96$.

Quantile function



Universality of the uniform distribution

- The c.d.f. is also a function, thus we can apply it to X itself.
- We can show that $U = F_X(X) \sim Unif(0,1)$:

$$\begin{aligned}\mathbb{P}(U \leq u) &= \mathbb{P}(F_X(X) \leq u) = \mathbb{P}(X \leq F^{-1}(u)) \\ &= F_X(F^{-1}(u)) = u.\end{aligned}$$

- This is the theoretical foundation of all randomization-based tests.
- Therefore, we can construct any random variable X from the uniform distribution:

$$X = F^{-1}(U), U \sim Unif(0,1).$$

- It is quite useful in simulation.

Universality of the uniform distribution