STA2001 Probability and Statistics (I)

Lecture 9

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review

- Negative binomial distribution with parameter p and r:

  X, the number of Bernoulli trials at which the rth success is observed, and its pmf takes the form of

$$\text{pmf: } f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \; x \in \overline{S} = \{r, r+1, \cdots\}$$

- Poisson distribution with parameter $\lambda > 0$:

  X, the number of occurrences of an event in <u>a unit interval</u> and its pmf takes the form of

$$\text{pmf: } f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \; x \in \overline{S} = \{0, 1, \cdots\}$$

Chapter 3 Continuous Distribution

Section 3.1 Random Variable of Continuous Type

Continuous RV

Recall that a RV $X : S \to \overline{S}$ is called a discrete RV if $\overline{S}$ contains finite or countably infinite number of outcomes.

Now we consider RVs with $\overline{S}$ that is an interval or unions of intervals, which are quite common (e.g., velocity of a vehicle traveling along the high way)

Discrete RV vs. Continuous RV

RV X is a function $X : S \to \overline{S} \subseteq R$

Discrete RV:

Continuous RV:

pmf $f(x) : \overline{S} \to (0,\ 1]$

1. $f(x) > 0$
2. $\sum_{x \in \overline{S}} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x)$

Continuous RV

Definition

A RV X with $\overline{S}$ that is an interval or unions of intervals is said to be continuous RV, if there exists a function $f(x)$:$\overline{S} \to (0, \infty)$ such that

1. $f(x) > 0, \quad x \in \overline{S}$
2. $\int_{\overline{S}} f(x)dx = 1$
3. If $[a, b] \subseteq \overline{S}$

$$P(a \leq X \leq b) \triangleq \int_a^b f(x)dx$$

f is the so called probability density function (pdf).

Discrete RV vs. Continuous RV

RV X is a function $X : S \to \overline{S} \subseteq R$

Discrete RV:

pmf $f(x) : \overline{S} \to (0,\ 1]$

1. $f(x) > 0$
2. $\sum_{x \in \overline{S}} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x)$

Continuous RV:

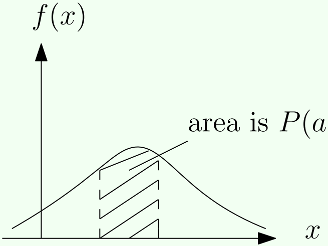
pdf $f(x) : \overline{S} \to (0, \infty)$

1. $f(x) > 0$
2. $\int_{\overline{S}} f(x) dx = 1$
3. $P(X \in A) = \int_A f(x) dx$

Interpretation of pdf

Interpretation

1. $$P(a \leq X \leq b) = \int_a^b f(x)dx$$

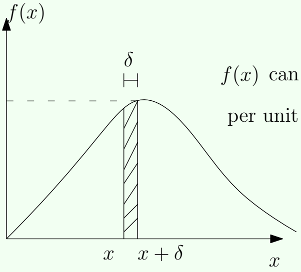


Interpretation of pdf

Interpretation

2.

$$P(x \le X \le x + \delta) = \int_x^{x+\delta} f(t)dt \approx f(x)\delta$$



$f(x)$ can be viewed as the probability mass per unit length near x

Remarks

1. We often extend the domain of $f(x)$ from $\overline{S}$ to R and let $f(x) = 0$, $x \notin \overline{S}$. In this case, $f(x) : R \to [0, \infty)$ and $\overline{S}$ is called the support of X.

Remarks

1. We often extend the domain of $f(x)$ from $\overline{S}$ to R and let

   $f(x) = 0,\ x \notin \overline{S}$. In this case, $f(x) : R \to [0, \infty)$ and $\overline{S}$ is

   called the support of X.

$$\begin{cases} f(x) \geq 0, \quad x \in R \\ \int_{-\infty}^{\infty} f(x)dx = 1 \\ P(a \leq X \leq b) = \int_a^b f(x)dx \end{cases}$$

Remarks

2. For any single value a, $P(X = a) = \int_a^a f(x)dx = 0$.

   Therefore, including or excluding the end points of an interval has no effect on its probability:

   $$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

Remarks

2. For any single value a, $P(X = a) = \int_a^a f(x)dx = 0$.

   Therefore, including or excluding the end points of an interval has no effect on its probability:

   $$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

3. pdf needs not to be continuous

   $$f(x) = \begin{cases} \frac{1}{2}, & 0 < x < 1, \quad 2 < x \leq 3 \\ 0, & \text{otherwise} \end{cases}$$

4. pdf needs not to be bounded, e.g., the Gamma distribution

Cumulative distribution function

Definition

cdf $F(x) : R \to [0,1]$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

1. $F(x)$ is nondecreasing
2. relation between the probability function and the cdf

$$P(a \le X \le b) = F(b) - F(a)$$

3. relation between the pdf and the cdf

$$f(x) = F'(x)$$

   for those values of x at which $F(x)$ is differentiable

Example 1 [Uniform Distribution]

Let the RV X denote the outcome when a point is selected randomly from $[a, b]$ with $-\infty < a < b < \infty$.

Define the pdf of X

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

What is the cdf of X?

Example 1 [Uniform Distribution]

Let the RV X denote the outcome when a point is selected randomly from $[a, b]$ with $-\infty < a < b < \infty$.

Define the pdf of X

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

What is the cdf of X?

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

Uniform Distribution

$$\text{For any } x \in [a, b], \qquad P(X \leq x) = \frac{x - a}{b - a}$$

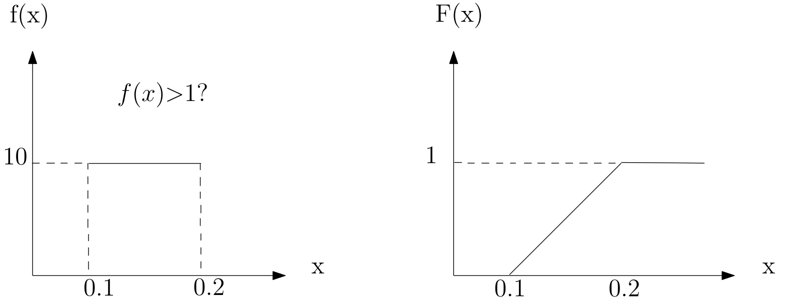
implies the probability of selecting a point from the interval $[a, x]$ is proportional to the length of $[a, x]$. Such distribution is called uniform distribution and denoted by $X \sim U(a, b)$.

Uniform Distribution

$$\text{For any } x \in [a, b], \qquad P(X \le x) = \frac{x - a}{b - a}$$

implies the probability of selecting a point from the interval $[a, x]$ is proportional to the length of $[a, x]$. Such distribution is called uniform distribution and denoted by $X \sim U(a, b)$.

For example, let $X \sim U(0.1, 0.2)$



Example 2, page 96

Let Y be a continuous RV with pdf $g(y) = 2y$, $0 < y < 1$.

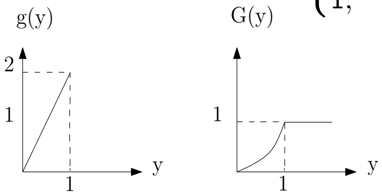
What is the cdf of Y, $P(\frac{1}{2} < Y \leq \frac{3}{4})$, $P(\frac{1}{4} < Y < 2)$?

Example 2, page 96

Let Y be a continuous RV with pdf $g(y) = 2y$, $0 < y < 1$.

What is the cdf of Y, $P(\frac{1}{2} < Y \leq \frac{3}{4})$, $P(\frac{1}{4} < Y < 2)$?

$$G(y) = P(Y \leq y) = \int_{-\infty}^{y} g(t)dt = \begin{cases} 0, & y \leq 0 \\ y^2, & 0 < y < 1 \\ 1, & y \geq 1 \end{cases}$$

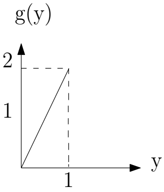


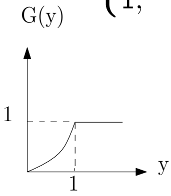
Example 2, page 96

Let Y be a continuous RV with pdf $g(y) = 2y$, $0 < y < 1$.

What is the cdf of Y, $P(\frac{1}{2} < Y \le \frac{3}{4})$, $P(\frac{1}{4} < Y < 2)$?

$$G(y) = P(Y \le y) = \int_{-\infty}^{y} g(t)dt = \begin{cases} 0, & y \le 0 \\ y^2, & 0 < y < 1 \\ 1, & y \ge 1 \end{cases}$$





$$P(\frac{1}{2} < Y \le \frac{3}{4}) = G(\frac{3}{4}) - G(\frac{1}{2}) = \frac{5}{16}$$

$$P(\frac{1}{4} < Y < 2) = G(2) - G(\frac{1}{4}) = \frac{15}{16}$$