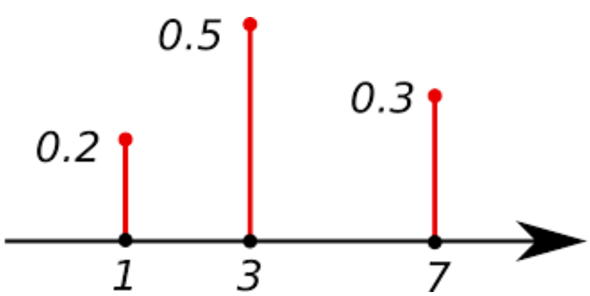
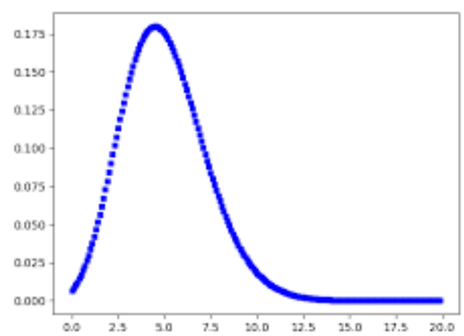
SPC

LESSON: Quality Methods - Probability Distributions

Quality Methods

Probability Distributions





Dr. Diane Evans
ROSE-HULMAN INSTITUTE OF TECHNOLOGY

Probability Distributions model different types of situations:

Binomial: Number of successes out of n identical trials in which each trial is either a success or failure; e.g., number of basketball foul shots made out of $n = 10$ shots

Exponential: Lifetimes or survival times of certain types of products; e.g., the lifetime of a spark plug

Geometric: Number of trials until the first success in identical trials; e.g., number of basketball foul shots until the first shot is made

Negative Binomial: Number of trials until the r^{th} success in identical trials; e.g., number of basketball foul shots until the r^{th} shot is made

Uniform: Lifetimes that are "uniformly" or randomly distributed over some interval; e.g., the wait time for a bus that comes every 20 mins

Poisson: Number of occurrences of an event over a given time period; e.g., number of customer arrivals at Java City in a 5 minute period

Random Variables: X, Y, Z

Continuous RV

- Takes on values on a continuous interval
- **Example:** X is a random variable with support $0 < x < 2$, or $x > 100$
- Described as **measurement** data
- **Examples:** height, weight, length of time, temperature
- Determine probabilities by finding areas under the curve of the function f(x)

Continuous RV

- Takes on "discrete" values; described as **count** data
- Finite or infinitely countable set of outcomes
- **Example:** X is a random variable with support x = 0, 1, 2, 3, ..., or support x = 2, 3, 5.
- **Example:** Number of tails when flipping a coin, number of computer crashes in a week

Discrete or Continuous RV's?

Example 1: Let X denote the number of defective transistors in a box of 10 transistors.
Discrete or Continuous RV? Support of *X*?

Example 2: Let T represent the time taken to assemble a certain part on an assembly line.
Discrete or Continuous RV? Support of *T*?

Example 3: Let Y represent the number of paint chips on a randomly selected car door from the Rose parking lot. Discrete or Continuous RV? Support of *Y*?

Example 4: Determine if each random variable is continuous or discrete.

(a) A student tracked the number of nervous tics he had per day during Spring Quarter 2010.

(b) The time required for a mailperson to finish their typical route.

(c) Number of medication errors tracked by customers at a given pharmacy.

(d) Length of the longest pencil in your book bag.

(e) Number of potato chips in a bag.

(f) Number of times my dogs wake me up at night.

(g) Weight of the heaviest book bag in our class.

Probability Mass Function or Probability Density Functions (PDFs)

Discrete RV **Continuous RV**

- Probability **Mass** Function: $p(x) = P(X = x)$
- Probability **Density** Function: $f(x)$

Properties of each:

- Both positive functions: $f(x) \geq 0$, $p(x) \geq 0$ for all x
- To determine probabilities for Discrete Random Variables, we SUM $p(x)$; e.g., $p(0) + p(1) + p(2)$
- To determine probabilities for Continuous Random Variables, we INTEGRATE $f(x)$; e.g., $\int_0^\infty f(x)dx$

Definition: The probabilities associated with a **Discrete Random Variable *X*** are specified by its **probability mass function *p(x)***. The function *p(x)* satisfies the following two properties:

$$(1)\ \ p(x) \geq 0 \text{ for all } x\text{, and}$$

$$(2)\ \ \sum_{\text{support of } X} p(x) = 1$$

The probability that the random variable X takes on values between (and including) x = *a* to x = *b* is:

$$p(a) + p(a+1) + p(a+2) + \ldots + p(b).$$

Example 5. Let X denote the number of customerswaiting in line at a Café at 10 a.m. on Monday. The probabilities associated with X are given by the following probability mass function:

x	0	1	2	3	4	5
$p(x)$	0.03	0.18	0.24	0.28	0.10	0.17

(a) Is $p(x)$ a legitimate probability mass function? Why or why not? Yes

$0.03 + 0.18 + 0.24 + 0.28 + 0.1 + 0.17 = 1$

(a) What's the probability that 3 or more customers are waiting for coffee at 10 a.m. on Monday?

$P(X \geq 3) = 0.28 + 0.10 + 0.17 = 0.55$

Definition: A **probability density function *f(x)*** is used to determine probabilities associated with a **continuous random variable *X***. The function f(x) satisfies the following two properties:

$$(1)\ \ f(x) \geq 0 \text{ for all } x\text{, and}$$

$$(2)\ \ \int_{\text{support of } X} f(x)dx = 1 \ \ \text{(i.e., area under } f(x) \text{ is 1)}$$

The probability that the random variable X takes on values between $x = a$ and $x = b$ is: $\int_{x=a}^{x=b} f(x)dx$

Example 6. Suppose I take a bus to work every day and a **bus arrives** at my bus stop **every 5 minutes**. Because I don't always leave my house at exactly the same time, I don't always arrive at the bus stop at the same time. Let **X represent my waiting time** (in minutes) at the bus stop. Then X is a continuous RV with support $0 \le x \le 5$. One possible probability density function that I can use to model X is:

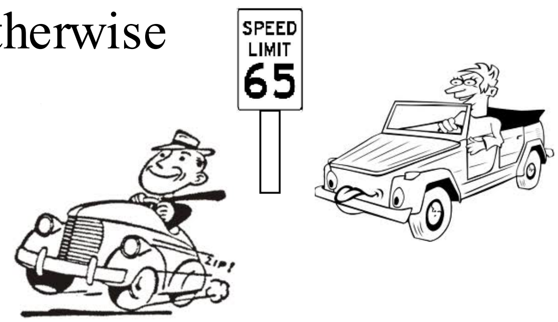
$$f(x) = \begin{cases} 1/5 & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

Is $f(x)$ a legitimate probability density function? Why or why not? Yes

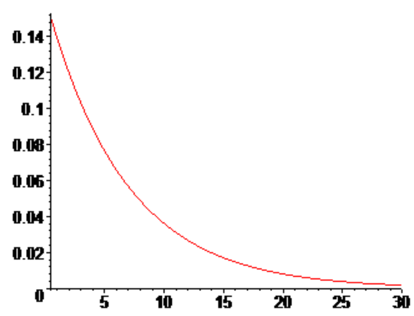
$f(x) \ge 0$, Area $= 1$ $\frac{1}{5}$ 5 $\int_0^5 \frac{1}{5}\,dx = \frac{1}{5}x\Big|_0^5 = 1$

Example 7. "Time headway" *X* in traffic flow is the elapsed time between one car passing a fixed point and the instant that the next car passes that same point. The probability density function *f(x)* is essentially the time headway (in seconds) for two randomly chosen consecutive cars on a freeway during a period of heavy flow.

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)} & x \ge 0.5 \\ 0 & \text{otherwise} \end{cases}$$



A graph of $f(x)$:



Is $f(x)$ a legitimate density function? YES! Why?

(1) $f(x) \ge 0$ for all real numbers x

(2) The area under $f(x)$ is 1: $\int_{0.5}^{\infty} 0.15e^{-0.15(x-0.5)}\,dx = 1$

Proportion of headway times between 0.5 and 5 seconds

$= \int_{0.5}^{5} 0.15e^{-0.15(x-0.5)}\,dx \cong 0.4908$

Expected Value E(X) or Mean of a RV

If X is a Discrete RV, its expected value or mean is given by:

$$\mu = E(X) = \sum_{\text{support of } X} p(x) \cdot x$$

If X is a Continuous RV, its expected value or mean is given by:

$$\mu = E(X) = \int_{-\infty}^{\infty} f(x) \cdot x \; dx$$

Example 5. (continued) Let *X* denote the number of customers waiting in line at a Café at 10 a.m. on Monday. The probability mass function of *X* is:

x	0	1	2	3	4	5
$p(x)$	0.03	0.18	0.24	0.28	0.10	0.17

What is the expected number of customers at this time?

$0(0.03) + 1(0.18) + 2(0.24) + 3(0.28)$
$+ 4(0.10) + 5(0.17) = 2.75$ cust

don't round!

Return to Example 6. Let ***X* represent my waiting time** (in minutes) at the bus stop. Recall that *X* is being modeled by the following probability density function:

$$f(x) = \begin{cases} 1/5 & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

What is my expected wait time at the bus stop?

$$E(X) = \int_0^5 \frac{1}{5} \cdot x \, dx = \frac{1}{10} x^2 \Big|_0^5 = \frac{25}{10} = 2.5 \text{ min}$$

Variance of a Discrete Random Variable

If *X* is a Discrete RV, its **variance** is given by:

$$\boldsymbol{\sigma^2 = \mathrm{Var}(X) = \sum_{\text{supp of } X} p(x) \cdot (x - \mu)^2}$$

or

$$\sigma^2 = \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

where

$$E(X^2) = \sum_{\text{supp of } X} p(x) \cdot x^2$$

It's the same formula for continuous RV's, except summations are replaced by integrals.

Return to Example 5. Let *X* denote the number of customers waiting in line at a Café at 10 a.m. on Monday. The probability mass function of *X* is:

$+\sqrt{1.9475} \simeq 1.396$ cust

x	0	1	2	3	4	5
$p(x)$	0.03	0.18	0.24	0.28	0.10	0.17

What is the variance of the number of customers at this time? How about standard deviation, where the standard deviation is the positive square root of the variance?

$E(X^2) - (E(X))^2 =$

$(0^2 \cdot 0.03 + 1^2 \cdot 0.18 + 2^2 \cdot 0.24 + 3^2 \cdot 0.28 + 4^2 \cdot 0.10 + 5^2 \cdot 0.17) - (2.75)^2 \simeq 1.9475 \text{ cust}^2$

Return to Example 6. Let ***X* represent my waiting time** (in minutes) at the bus stop. Recall that X is being modeled by the following probability density function:

$$f(x) = \begin{cases} 1/5 & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

What is the variance of my wait time at the bus stop? How about standard deviation?

$E(X^2) - (E(X))^2 = \int_0^5 \frac{1}{5} \cdot x^2 \, dx - (2.5)^2$

$= \frac{1}{15} x^3 \Big|_0^5 - (2.5)^2 \simeq 2.08 \text{ mi}^2 ; \ \sigma \approx 1.44 \text{ mi}$