AFRE 802
Statistical Methods for Agricultural, Food, & Resource Economists

Continuous random variables & their probability distributions – Part 1 of 3 (WMS Ch. 4.1-4.3)

September 26, 2017

Nicole Mason
Michigan State University
Fall 2017

GAME PLAN

- Collect Ch. 3 HW
- Reminder: Integration/C&W Ch. 14 HW due Thursday
- Review
- Graded in-class exercise
- Probability distributions for continuous RVs
  1. Cumulative distribution functions (CDFs)
  2. Probability density functions (PDFs)
  3. Expected values & variances
- Next class: specific, common continuous probability distributions

Review

MICHIGAN STATE UNIVERSITY

Tchebysheff's Inequality

For **any** RV, *Y,* with with mean μ & variance σ^2:

$$P\left[(\mu-k\sigma)<Y<(\mu+k\sigma)\right]\geq 1-\frac{1}{k^2}$$

or

$$P\left[Y\leq(\mu-k\sigma)\text{ OR }Y\geq(\mu+k\sigma)\right]\leq\frac{1}{k^2}$$

for any constant $k>0$

The probability of being less than k standard deviations from the mean is at least $1-1/k^2$

The probability of being k or more standard deviations from the mean is no more than $1/k^2$

Review

$$\frac{dF(x)}{dx}=f(x)\quad\Rightarrow\quad\int f(x)\,dx=F(x)+c$$

MICHIGAN STATE UNIVERSITY

Integration: the reverse of differentiation

- Definite vs. indefinite integrals
- Definite integrals as the area under f(x)
- Main rules:

1. Power rule: $\int x^n\,dx=\frac{x^{n+1}}{n+1}+c\qquad(n\neq-1)$
2. Exponential rule: $\int e^x\,dx=e^x+c$ and $\int f'(x)e^{f(x)}\,dx=e^{f(x)}+c$
3. Logarithmic rule: $\int\frac{1}{x}dx=\ln x+c\ \ (x>0)$ and $\int\frac{f'(x)}{f(x)}dx=\ln f(x)+c\ \ (f(x)>0)$
4. Integral is a linear operator: $\int\left[af(x)+bg(x)\right]dx=a\int f(x)\,dx+b\int g(x)\,dx$
5. Substition rule: $\int f(u)\frac{du}{dx}dx=\int f(u)\,du=F(u)+c$
6. Integration by parts: $\int f'(x)g(x)\,dx=f(x)g(x)-\int f(x)g'(x)\,dx$

- See lecture notes from last class for properties of definite integrals
- *Questions from last class?*

MICHIGAN STATE UNIVERSITY

Graded in-class exercise

MICHIGAN STATE UNIVERSITY

Continuous RV & their probability distributions

Discrete vs. continuous RVs

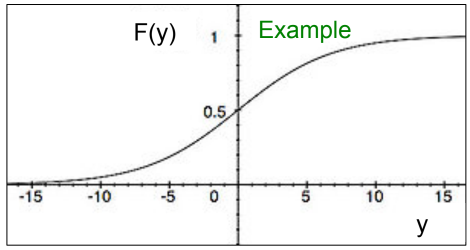
- **Discrete**: an RV that takes on a finite or countably infinite # of distinct values
- **Continuous:** *definition (# of outcomes & probability of each outcome)?*
  - An RV that takes on an uncountably infinite # of values (in its range)
    - *Examples?*
  - $P(Y=y)=0$, i.e., the probability of an individual value of y is zero (infinite # of possible values so can't count or assign a positive probability to them)

The probability distribution of a continuous RV

- *How did we specify the probability distribution for discrete RVs?*
  - Assign a probability to each distinct value that the discrete RV could take on; probabilities in [0,1] and sum to 1
- This approach isn't possible w/ continuous RVs b/c we have an infinite # of values
- Instead, we'll specify a distribution by its:
  - **Cumulative distribution function (CDF)**, and
  - **Probability density function (PDF)**

Cumulative distribution function (CDF)

$$\boxed{F(y) = P(Y \le y) \text{ for } -\infty \le y \le \infty}$$



Properties

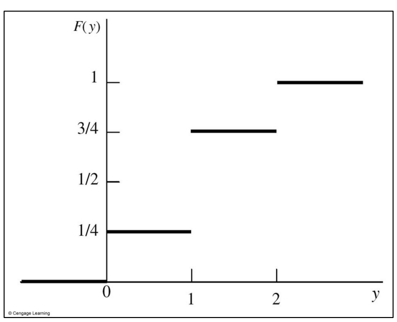
1. $F(-\infty) \equiv \lim_{y \to -\infty} F(y) = 0$

2. $F(\infty) \equiv \lim_{y \to \infty} F(y) = 1$

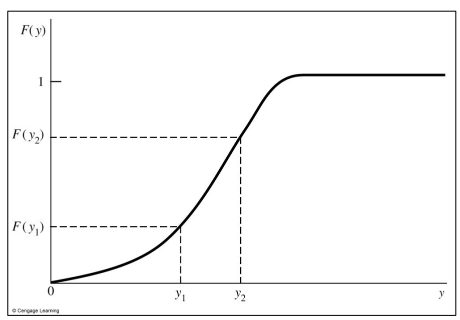
3. $F(y)$ is a nondecreasing function of y.

[so if $y_1 < y_2$, $F(y_1) \le F(y_2)$]

Examples of CDFs for discrete vs. continuous RVs



Discrete RV



Continuous RV

- Check that both satisfy the 3 properties of CDFs
- *Why is the discrete CDF a "step function", whereas the continuous CDF is smooth?*
- An RV, *Y*, is said to be **continuous if its CDF, *F(y)*, is continuous** for $-\infty < y < \infty$

Probability density function (PDF)

= an equation to describe the probability distribution of a continuous RV

$$\boxed{f(y) = \frac{dF(y)}{dy} = F'(y)}$$

Properties of PDFs for continuous RVs:

1. $f(y) \geq 0$ for all y, $-\infty < y < \infty$

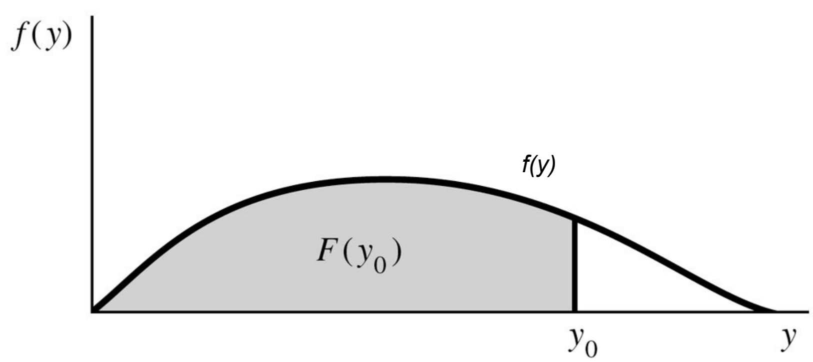
2. $\int_{-\infty}^{\infty} f(y)\,dy = 1$

Look familiar? Analogues for discrete RVs?

Relationships b/w CDFs & PDFs

$$\boxed{f(y) = \frac{dF(y)}{dy} = F'(y)}$$

$$\boxed{F(y) = \int_{-\infty}^{y} f(t)\,dt}$$



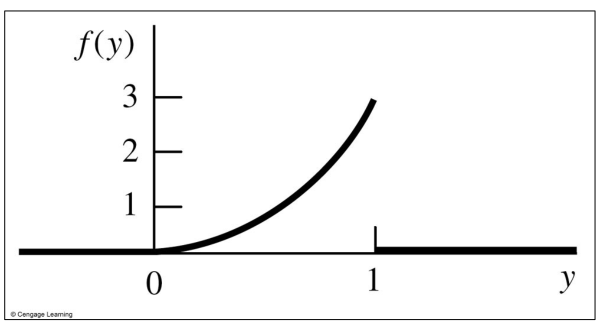
PDFs & CDFs – example #1

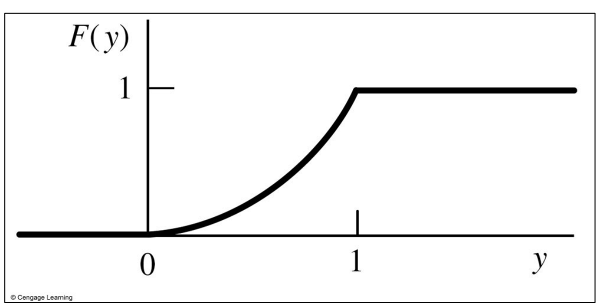
$$F(y) = \int_{-\infty}^{y} f(t)\, dt$$

Suppose the PDF for continuous RV, Y, is:

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

Find $F(y)$, then graph both $f(y)$ and $F(y)$.





PDFs & CDFs – example #2

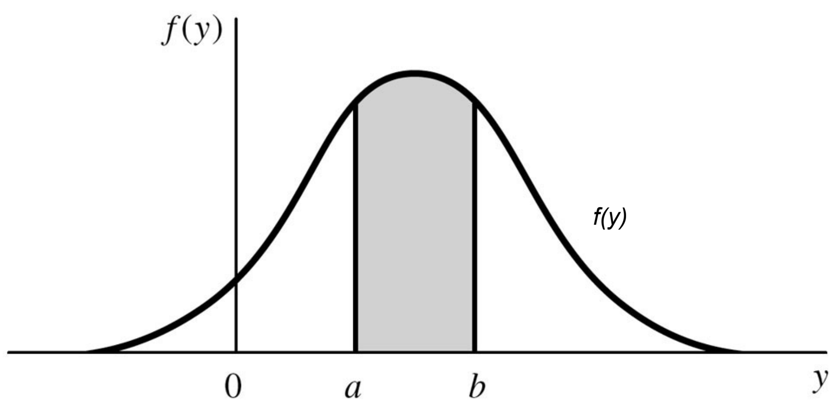
Properties of PDFs for continuous RVs:

1. $f(y) \geq 0$ for all y, $-\infty < y < \infty$
2. $\int_{-\infty}^{\infty} f(y)\,dy = 1$

Suppose $f(y)=cy^2$ for $0 \leq y \leq 2$, and $f(y)=0$ elsewhere. Find the value of c for which $f(y)$ is a valid density function.

MICHIGAN STATE UNIVERSITY

How can we find $P(a \leq Y \leq b)$? (use the figure below)



$$P(a \leq Y \leq b) = \int_a^b f(y)\,dy = F(b) - F(a)$$

Recall that *P(Y=y)=0* for a continuous RV

What does this imply about the probabilities below?

$$P(a \le Y \le b)$$
$$P(a \le Y < b)$$
$$P(a < Y \le b)$$
$$P(a < Y < b)$$

All are equal!

$$P(a \le Y \le b) = \int_a^b f(y)\, dy = F(b) - F(a)$$

PDFs & CDFs – example #3

Find $P(1 \le Y \le 2)$ and $P(1 < Y < 2)$ for the RV in example #2. Recall that the PDF was $f(y)=(3/8)y^2$ for $0 \le y \le 2$, and $f(y)=0$ elsewhere.

MICHIGAN STATE UNIVERSITY

The expected value of a continuous RV

- Recall the formula for *E(Y)* for **discrete** *Y*:

$$E(Y) = \sum_i y_i p(y_i)$$

- *What is the formula for the expected value of a continuous RV (i.e., what is the continuous analogue of the expression above)?*

$$\boxed{E(Y) = \int_{-\infty}^{\infty} y f(y)\, dy}$$

Expected value of a continuous RV

MICHIGAN STATE UNIVERSITY

The expected value of a function of a continuous RV

- Recall the formula for *E[g(Y)]* for **discrete** *Y*:

$$E[g(Y)] = \sum_i g(y_i) p(y_i)$$

- *What is the formula for the expected value of a function of a continuous RV (i.e., what is the continuous analogue of the expression above)?*

$$\boxed{E[g(Y)] = \int_{-\infty}^{\infty} g(y) f(y)\, dy}$$

Expected value of a function of a continuous RV

MICHIGAN STATE UNIVERSITY

All of the same **expected value rules** that applied to discrete RVs apply to continuous RVs

For any constants b and c:

$(i)\ E(c) = c$

$(ii)\ E(bX) = bE(X)$

$(iii)\ E(bX + c) = bE(X) + c$

$$(iv)\ E\left[g_1(X) + g_2(X) + \ldots + g_k(X)\right] = E\left[g_1(X)\right] + E\left[g_2(X)\right] + \ldots + E\left[g_k(X)\right]$$

MICHIGAN STATE UNIVERSITY

Likewise for **variance** of a continuous RV

$$\boxed{V(Y) = E\left[(Y-\mu)^2\right] = E(Y^2) - \mu^2 = E(Y^2) - \left[E(Y)\right]^2}$$

Variance of an RV (discrete or cont.)

$$\boxed{V(Y) = \sum_i (y_i - \mu)^2 p(y_i) = \left[\sum_i y_i^2 p(y_i)\right] - \left[\sum_i y_i p(y_i)\right]^2}$$

Variance of a discrete RV

- *How would you express V(Y) for a continuous variable in terms of its PDF, f(y), and integrals?*

$$\boxed{V(Y) = \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\,dy = \int_{-\infty}^{\infty} y^2 f(y)\,dy - \left[\int_{-\infty}^{\infty} y f(y)\,dy\right]^2}$$

Variance of a continuous RV

For any constants b and c:

$(i)\ V(c) = 0$

$(ii)\ V(bX) = b^2 V(X)$

$(iii)\ V(bX + c) = b^2 V(X)$

MICHIGAN STATE UNIVERSITY

E(Y) and *V(Y)* for continuous RV - example

In example #2 earlier, we found $f(y)=(3/8)y^2$ for $0 \le y \le 2$, and $f(y)=0$ elsewhere. If random variable Y has this continuous PDF, find $E(Y)$ and $V(Y)$.

$$\boxed{E(Y) = \int_{-\infty}^{\infty} y f(y)\,dy}$$

$$\boxed{\begin{aligned} V(Y) &= E(Y^2) - [E(Y)]^2 \\ E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y)\,dy \end{aligned}}$$

MICHIGAN STATE UNIVERSITY

Homework:

- WMS Ch. 4 (part 1 of 3)
  - PDFs & CDFs: 4.8, 4.11
  - Expected values & variances of continuous RVs: 4.20, 4.25, 4.33 (parts a & c only)
- Reminder: C&W Ch. 14 HW (integration) due on Thursday

Next class:

- Continuous random variables (Part 2 of 3) – specific distributions (uniform, normal)

Reading for next class:

- WMS Ch. 4 (sections 4.4-4.5)

Application for next class:

- Look up application in your field of uniform or normal distribution

In-class exercises #1 – CDF & probabilities for a continuous RV

4.13 A supplier of kerosene has a 150-gallon tank that is filled at the beginning of each week. His weekly demand shows a relative frequency behavior that increases steadily up to 100 gallons and then levels off between 100 and 150 gallons. If Y denotes weekly demand in hundreds of gallons, the relative frequency of demand can be modeled by

$$f(y) = \begin{cases} y, & 0 \le y \le 1, \\ 1, & 1 < y \le 1.5, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find $F(y)$.
b Find $P(0 \le Y \le .5)$.
c Find $P(.5 \le Y \le 1.2)$.

In-class exercises #2 – *E(Y)* & *V(Y)* for continuous RV

4.21 If, as in Exercise 4.17, Y has density function

$$f(y) = \begin{cases} (3/2)y^2 + y, & 0 \le y \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

find the mean and variance of Y.

MICHIGAN STATE UNIVERSITY

In-class exercises #3 – *E(Y)* & *V(Y)* for continuous RV

4.32 Weekly CPU time used by an accounting firm has probability density function (measured in hours) given by

$$f(y) = \begin{cases} (3/64)y^2(4-y), & 0 \le y \le 4, \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the expected value and variance of weekly CPU time.

b The CPU time costs the firm $200 per hour. Find the expected value and variance of the weekly cost for CPU time.

c Would you expect the weekly cost to exceed $600 very often? Why?