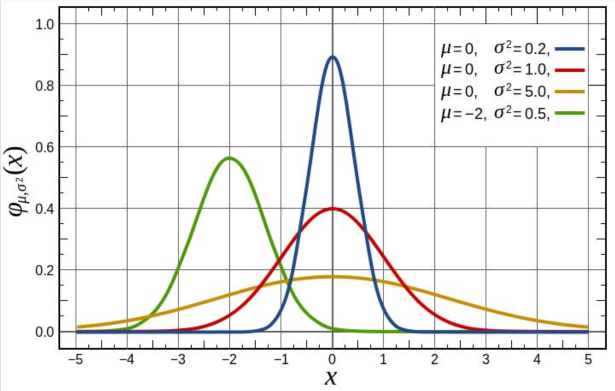
Variance; Continuous Random Variables
18.05 Spring 2022



Announcements/Agenda

Announcements

- None

Agenda

- Studio 2 comments
- Variance and standard deviation for discrete variables
- Calculus warmup
- Continuous random variables - probability densities
- Exponential distribution

Studio 2 comments

- Graded studio 2 code is posted in the usual place.
- Excellent job overall!
- Please suppress stray printouts. We will start penalizing these.
- If you used loops, please look at the solutions to see how R lets you do array operations without loops.
- Expected value of payoff IS NOT payoff of expected value.

  Example. $X \sim \text{Bernoulli}(p)$ and $Y = X^2 + X$.

  $$E[X] = p, \quad E[Y] = 2p, \quad E[X]^2 + E[X] = p^2 + p \neq E[Y].$$

Variance and standard deviation

X a discrete random variable with mean $E[X] = \mu$.

- Meaning: spread of probability mass about the mean.
- Definition as expectation (weighted sum):

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

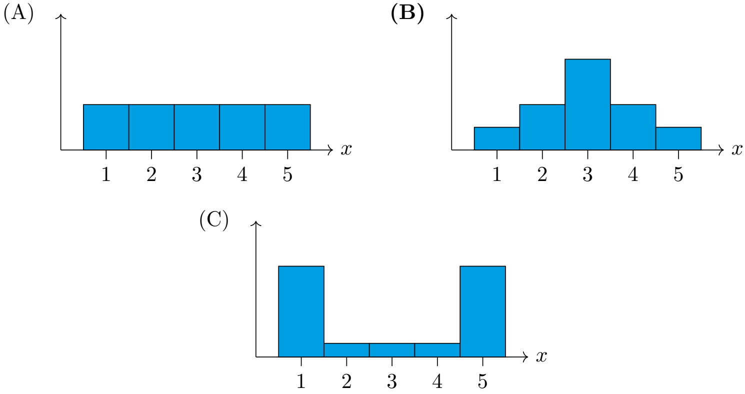
- Computation as sum:

$$\text{Var}(X) = \sum_{i=1}^{n} p(x_i)(x_i - \mu)^2.$$

- Standard deviation $\sigma = \sqrt{\text{Var}(X)}$.
  Units for standard deviation = units of X.

1. Concept question: Order the variance

The graphs below give the pmf for 3 random variables.



Order them by size of standard deviation from biggest to smallest.
(Assume x has the same units in all three.)

1. ABC 2. ACB 3. BAC 4. BCA 5. CAB 6. CBA

2. Concept question: Zero variance

Suppose X is a discrete random variable,

True or False: If $\text{Var}(X) = 0$ then X is constant.

1. True
2. False

Computation from tables

Example. Compute the variance and standard deviation of X.

values x	1	2	3	4	5
pmf $p(x)$	$1/10$	$2/10$	$4/10$	$2/10$	$1/10$

Computation from tables

Example. Compute the variance and standard deviation of X.

values x	1	2	3	4	5
pmf $p(x)$	1/10	2/10	4/10	2/10	1/10

A very useful formula

The following formula is often easier to use than the definition.

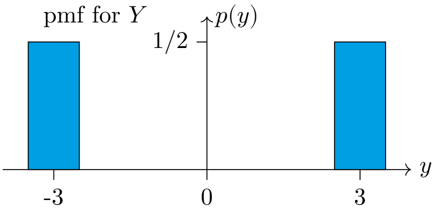
$$\text{Var}(X) = E[X^2] - E[X]^2 = \left(\sum_{i=1}^{n} p(x_i)x_i^2\right) - \mu^2.$$

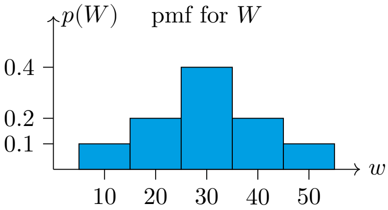
Redo the above computation using this formula.

(Written solution with posted solutions)

3. Concept question: Standard deviation

Make an intuitive guess: Which pmf has the bigger standard deviation? (Assume w and y have the same units.)

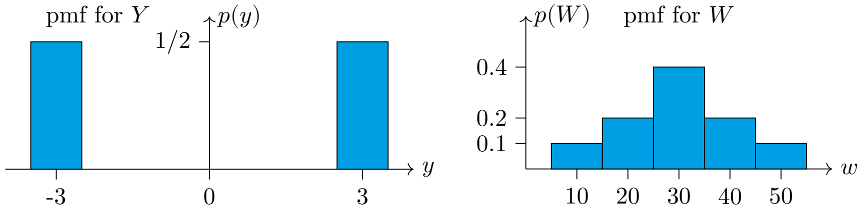




1. Y 2. W

3. Concept question: Standard deviation

Make an intuitive guess: Which pmf has the bigger standard deviation? (Assume w and y have the same units.)



1. Y 2. W

Table question: make probability tables for Y and W and compute their standard deviations.

Algebraic properties of variance

If a and b are constants then

$$\mathsf{Var}(aX + b) = a^2\,\mathsf{Var}(X), \qquad \sigma_{aX+b} = |a|\,\sigma_X.$$

If X and Y are **independent** random variables then

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y).$$

Board questions

(a) Let $X \sim \text{Bernoulli}(p)$. Compute $\text{Var}(X)$.

(b) Let $Y \sim \text{Bin}(n, p)$. Show $\text{Var}(Y) = n\,p(1-p)$.

(c) Suppose $X_1, X_2, \ldots, X_n$ are independent and all have the same standard deviation $\sigma = 2$. Let $\overline{X}$ be the average of $X_1, \ldots, X_n$.

What is the standard deviation of $\overline{X}$?

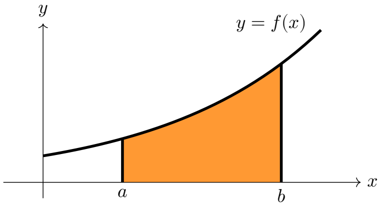
Continuous random variables

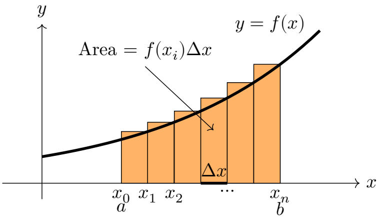
- Like discrete, except take a continuous range of values
- Replace probability mass function by probability density function
- Replace sums by integrals

Calculus warmup for continuous random variables

1. $\int_a^b f(x)\,dx =$ area under the curve $y = f(x)$.
2. $\int_a^b f(x)\,dx =$ 'sum of $f(x)\,dx$'.

Connection between the two views:





Area is approximately the sum of rectangles:

$$\int_a^b f(x)\,dx \approx f(x_1)\Delta x + f(x_2)\Delta x + \ldots + f(x_n)\Delta x = \sum_1^n f(x_i)\Delta x.$$

Continuous random variables: pdf and cdf

- Continuous range of values:

$$[0,1],\ [a,b],\ [0,\infty),\ (-\infty,\infty).$$

- Probability density function (pdf)

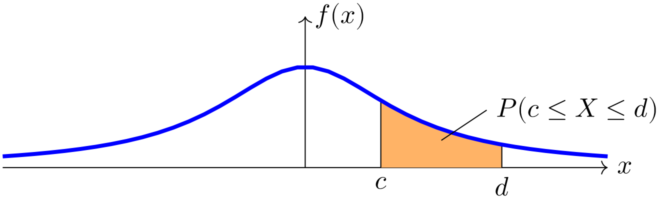
$$f(x) \geq 0; \quad P(c \leq X \leq d) = \int_c^d f(x)\, dx = \text{ 'sum' of } f(x)dx.$$

Units for the pdf are $\dfrac{\text{prob.}}{\text{unit of x}}$ (This explains the term density.)

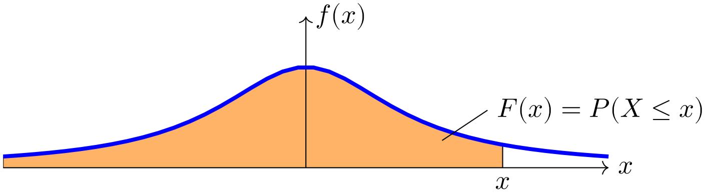
- Cumulative distribution function (cdf)

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

Visualization



pdf and probability



pdf and cdf

Properties of the cdf

(Same as for discrete distributions)

- (Definition) $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du$.
- $0 \leq F(x) \leq 1$.
- non-decreasing.
- 0 to the left: $\lim_{x \to -\infty} F(x) = 0$.
- 1 to the right: $\lim_{x \to \infty} F(x) = 1$.
- $P(c < X \leq d) = F(d) - F(c)$.
- $F^{'}(x) = f(x)$.

Board questions

1. Suppose X has range $[0, 2]$ and pdf $f(x) = c\,x^2$.

(a) What is the value of c?

(b) Compute the cdf $F(x)$.

(c) Compute $P(1 \leq X \leq 2)$.

(d) Plot the pdf and use it to illustrate part (c).

2. Suppose Y has range $[0, b]$ and cdf $F(y) = y^2/9$.

(a) What is b?

(b) Find the pdf of Y.

4. Discussion questions

Suppose X is a continuous random variable.

(a) If the pdf of X is $f(x)$ can there be an x where $f(x) = 10$?

4. Discussion questions

Suppose X is a continuous random variable.

(a) If the pdf of X is $f(x)$ can there be an x where $f(x) = 10$?

(b) What is $P(X = a)$?

4. Discussion questions

Suppose X is a continuous random variable.

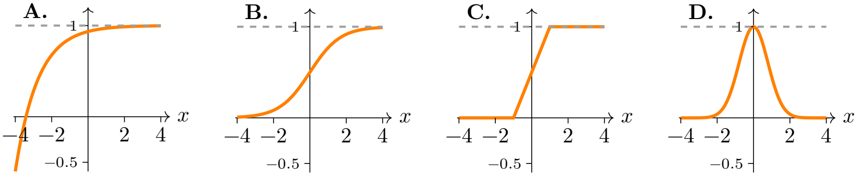
(a) If the pdf of X is $f(x)$ can there be an x where $f(x) = 10$?

(b) What is $P(X = a)$?

(c) Does $P(X = a) = 0$ mean X never equals a?

Discussion questions

Which of the following are graphs of valid cumulative distribution functions?



Exponential Random Variables

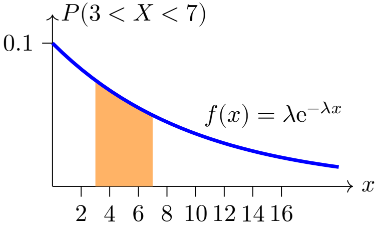
Parameter: λ (called the rate parameter).

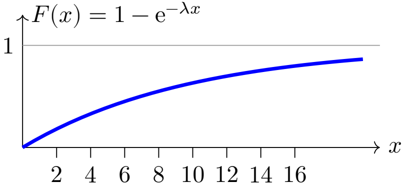
Range: $[0, \infty)$.

Notation: exponential(λ) or exp(λ).

Density: $f(x) = \lambda e^{-\lambda x}$ for $0 \leq x$.

Models: Waiting time





Continuous analogue of geometric distribution!

Board question

I've noticed that taxis drive past 77 Mass. Ave. on the average of once every 10 minutes.

Suppose time spent waiting for a taxi is modeled by an exponential random variable

$$X \sim \text{Exponential}(1/10); \qquad f(x) = \frac{1}{10}\mathrm{e}^{-x/10}$$

(a) Sketch the pdf of this distribution

(b) Shade the region which represents the probability of waiting between 3 and 7 minutes

(c) Compute the probability of waiting between between 3 and 7 minutes for a taxi

(d) Compute and sketch the cdf.