

Master Program in *Data Science and Business Informatics*

# Statistics for Data Science

Lesson 09 - Expectation and variance. Computations with random variables

Salvatore Ruggieri

Department of Computer Science

University of Pisa, Italy

[salvatore.ruggieri@unipi.it](mailto:salvatore.ruggieri@unipi.it)

# Expectation of a discrete random variable

- Buy lottery ticket every week,  $p = 1/10000$ , what is probability of winning at  $k^{\text{th}}$  week?

$$X \sim \text{Geo}(p) \quad P(X = k) = (1 - p)^{k-1} \cdot p \text{ for } k = 1, 2, \dots$$

- What is the average number of weeks to wait (expected) before winning?

$$E[X] = \sum_{k=1}^{\infty} k \cdot (1 - p)^{k-1} \cdot p = \frac{1}{p}$$

because  $\sum_{k=1}^{\infty} k \cdot x^{k-1} = 1/(1-x)^2$

DEFINITION. The *expectation* of a discrete random variable  $X$  taking the values  $a_1, a_2, \dots$  and with probability mass function  $p$  is the number

$$E[X] = \sum_i a_i P(X = a_i) = \sum_i a_i p(a_i).$$

- Expected value, mean value (weighted by probability of occurrence), center of gravity

See [seeing-theory.brown.edu](http://seeing-theory.brown.edu)

# Expected value may be infinite or may not exist!

- Fair coin: win  $2^k$  euros if first  $H$  appears at  $k^{\text{th}}$  toss [St. Petersburg paradox]
  - ▶  $X$  with p.m.f.  $p(2^k) = 2^{-k}$  for  $k = 1, 2, \dots$
  - ▶  $p(\cdot)$  is a p.m.f. since  $\sum_{k=1}^{\infty} 2^{-k} = 1$
  - ▶  $E[X] = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k} = \sum_{k=1}^{\infty} 1 = \infty$
- Expectation does not exist when  $\sum_i a_i p(a_i)$  does not converge
  - ▶  $X$  with p.m.f.  $p(2^k) = p(-2^k) = 2^{-k}$  for  $k = 2, 3, \dots$
  - ▶  $E[X] = \sum_{k=2}^{\infty} (2^k \cdot 2^{-k} - 2^k \cdot 2^{-k}) = \sum_{k=2}^{\infty} (1 - 1) = 0$  *wrong!*
  - ▶  $E[X] = \sum_{k=2}^{\infty} 2^k \cdot 2^{-k} - \sum_{k=2}^{\infty} 2^k \cdot 2^{-k} = \infty - \infty$  *undefined*
  - ▶  $E[X]$  is finite if  $\sum_i |a_i| p(a_i) < \infty$
  - ▶ In the case above,  $\sum_{k=2}^{\infty} (|2^k| \cdot 2^{-k} + |-2^k| \cdot 2^{-k}) = \infty$

using  $\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$  for  $|a| < 1$

# Expectation of some other discrete distributions

- Expectation of some other discrete distributions

- ▶  $X \sim U(m, M)$      $E[X] = (m+M)/2$

- $\sum_{i=m}^M \frac{i}{M-m+1} = \frac{1}{M-m+1} \sum_{i=0}^{M-m} (m+i) = m + (M-m)/2 = \frac{m+M}{2}$

- ▶  $X \sim \text{Ber}(p)$      $E[X] = p$

- $0 \cdot (1-p) + 1 \cdot p = p$

*[Mean may not belong to the support]*

- ▶  $X \sim \text{Bin}(n, p)$      $E[X] = n \cdot p$

- Because ... we'll see later

- ▶  $X \sim \text{NBin}(n, p)$      $E[X] = \frac{n \cdot p}{1-p}$

- Because ... we'll see later

- ▶  $X \sim \text{Poi}(\mu)$      $E[X] = \mu$

- Because, when  $n \rightarrow \infty$ :  $\text{Bin}(n, \mu/n) \rightarrow \text{Poi}(\mu)$

# Expectation of a continuous random variable

DEFINITION. The *expectation* of a continuous random variable  $X$  with probability density function  $f$  is the number

$$E[X] = \int_{-\infty}^{\infty} xf(x) dx.$$

- Expectation of some continuous distributions

- ▶  $X \sim U(\alpha, \beta)$      $E[X] = (\alpha + \beta)/2$

- ▶  $X \sim \text{Exp}(\lambda)$      $E[X] = 1/\lambda$

- Because  $\int_0^{\infty} x\lambda e^{-\lambda x} dx = [-e^{-\lambda x}(x + 1/\lambda)]_0^{\infty} = e^0(0 + 1/\lambda)$

[See Lesson 06]

- ▶  $X \sim N(\mu, \sigma^2)$      $E[X] = \mu$

- Because:  $\int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx = \mu + \int_{-\infty}^{\infty} (x - \mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx =_{z=\frac{x-\mu}{\sigma}} \mu + \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \mu$

- ▶  $X \sim \text{Erl}(n, \lambda)$      $E[X] = n/\lambda$

- Because ... we'll see later

# Expected value may not exist!

- Cauchy distribution

$$f(x) = \frac{1}{\pi(1+x^2)}$$

- $X_1, X_2 \sim N(0, 1)$  i.i.d.,  $X = X_1/X_2 \sim \text{Cau}(0, 1)$

$$E[X] = \int_{-\infty}^0 xf(x)dx + \int_0^{\infty} xf(x)dx$$

- $\int_{-\infty}^0 xf(x)dx = \left[\frac{1}{2\pi} \log(1+x^2)\right]_{-\infty}^0 = -\infty$

- $\int_0^{\infty} xf(x)dx = \left[\frac{1}{2\pi} \log(1+x^2)\right]_0^{\infty} = \infty$

$$E[X] = -\infty + \infty$$

- $E[X]$  is finite if  $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$

*Mean value does not always make sense in your data analytics project!*

# $E[g(X)] \neq g(E[X])$

- Recall that *velocity* = *space/time*, and then *time* = *space/velocity*!
- Vector  $v$  of speed (Km/h) to reach school and their probabilities  $p$  using feet, bike, bus, train:

$$v = c(5, 10, 20, 30) \quad p = c(0.1, 0.4, 0.25, 0.25)$$

- Distance house-schools is 2 Km
- What is the average time to reach school?
  - ▶  $2/\text{sum}(v*p)$  i.e.,  $\text{space}/E[\text{velocity}]$
  - ▶  $\text{sum}(2/v*p)$  i.e.,  $E[\text{space}/\text{velocity}]$
- $X = \text{velocity}$ ,  $g(X) = 2/X$  time to reach school
  - ▶  $E[g(X)] \neq g(E[X])$

# The change of variable formula (or rule of the lazy statistician)

- $X \sim U(0, 10)$ , width of a square field,  $E[X] = 5$
- $g(X) = X^2$  is the area of the field,  $E[g(X)] = ?$
- $F_g(a) = P(g(X) \leq a) = P(X \leq \sqrt{a}) = \sqrt{a}/10$  for  $0 \leq a \leq 100$
- Hence,  $f_g(a) = dF_g(a)/da = 1/20\sqrt{a}$
- $E[g(X)] = \frac{1}{20} \int_0^{100} \frac{x}{\sqrt{x}} dx = \frac{1}{20} \frac{2}{3} [x^{3/2}]_0^{100} = 100/3$
- Alternatively,  $E[g(X)] = \int_0^{10} x^2 \frac{1}{10} dx = \frac{1}{10} \frac{1}{3} [x^3]_0^{10} = 100/3$

$$[E[g(X)] \neq g(E[X])]$$

[later on, a general theorem]

**THE CHANGE-OF-VARIABLE FORMULA.** Let  $X$  be a random variable, and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a function.

If  $X$  is discrete, taking the values  $a_1, a_2, \dots$ , then

$$E[g(X)] = \sum_i g(a_i)P(X = a_i).$$

If  $X$  is continuous, with probability density function  $f$ , then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx.$$

See R script



## Theorem (Change of units)

$$E[rX + s] = rE[X] + s$$

- Example: for  $Y = 1.8X + 32$ , we have  $E[Y] = 1.8E[X] + 32$       [*Celsius to Fahrenheit*]

- **Corollary.**

$$E[X - E[X]] = E[X] - E[X] = 0$$

- **Theorem.** Expectation minimizes the square error, i.e., for  $a \in \mathbb{R}$ :

$$E[(X - E[X])^2] \leq E[(X - a)^2]$$

- ▶ Proof. (sketch) set  $\frac{d}{da} \int_{-\infty}^{\infty} (x - a)^2 f(x) dx = 0$

# Entropy of a random variable

- The **Shannon's information entropy** is the average level of “information”, “surprise”, or “uncertainty” inherent to the variable's possible outcomes
  - ▶ Information is inversely proportional to probability  $\frac{1}{p(a_i)}$ 
    - Highly likely events carry very little new information
    - Highly unlikely events carry more information
  - ▶ Information content  $ic()$  of two independent events should sum up  $\log \frac{1}{p(a_i)}$ 
    - $ic(p(A \cap B)) = ic(p(A)) + ic(p(B)) = ic(p(A)p(B))$
    - $ic(p(\Omega)) = ic(1) = 0$
    - $ic(p(A)) \geq 0$

- $H(X) = E[-\log p(X)]$  (*discrete*)       $H(X) = E[-\log f(X)]$  (*continuous*)

$$H(X) = - \sum_i p(a_i) \log p(a_i)$$

$$H(X) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx$$

- ▶ For  $X$  discrete,  $H(X) \geq 0$  since  $-\log p(X) = \log \frac{1}{p(X)} \geq 0$ 
  - reached when  $p(a_1) = 1$  and  $p(a_i) = 0$  for  $i \neq 1$
- ▶ For  $X \sim \text{Ber}(p)$ ,  $H(X) = -p \log p - (1-p) \log (1-p)$

See R script

# Computation with discrete random variables

## Theorem

For a discrete random variable  $X$ , the p.m.f. of  $Y = g(X)$  is:

$$P_Y(Y = y) = \sum_{g(x)=y} P_X(X = x) = \sum_{x \in g^{-1}(y)} P_X(X = x)$$

- **Proof.**  $\{Y = y\} = \{g(X) = y\} = \{x \in g^{-1}(y)\}$
- **Corollary** (the change-of-variable formula):

$$E[g(X)] = \sum_y y P_Y(Y = y) = \sum_y y \sum_{g(x)=y} P_X(X = x) = \sum_x g(x) P_X(X = x)$$

# Example

- $X \sim U(1, 200)$  number of tickets sold
- Capacity is 150
- $Y = \max\{X - 150, 0\}$  overbooked tickets

$$P_Y(Y = y) = \begin{cases} 150/200 & \text{if } y = 0 & g^{-1}(0) = \{1, \dots, 150\} \\ 1/200 & \text{if } 1 \leq y \leq 50 & g^{-1}(y) = \{y + 150\} \end{cases}$$

- Hence:

$$E[Y] = 0 \cdot \frac{150}{200} + \frac{1}{200} \cdot \sum_{y=1}^{50} y = 6.375$$

- or using the change-of-variable formula:

$$E[Y] = \frac{1}{200} \cdot \sum_{x=1}^{200} \max\{X - 150, 0\} = \frac{1}{200} \cdot \sum_{x=151}^{200} (X - 150) = 6.375$$

# Computation with continuous random variables

## Theorem

For a continuous random variable  $X$ , the density functions of  $Y = g(X)$  when  $g(\cdot)$  is increasing/decreasing are:

$$F_Y(y) = F_X(g^{-1}(y)) \quad f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

- **Proof.** (for  $g(\cdot)$  increasing) Since  $g(\cdot)$  is invertible and  $g(x) \leq y$  iff  $x \leq g^{-1}(y)$ :

$$F_Y(y) = P_Y(g(X) \leq y) = P_X(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

and then:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(g^{-1}(y))}{dy} = \frac{dF_X(g^{-1}(y))}{dg^{-1}} \frac{dg^{-1}(y)}{dy} = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}$$

**Exercise:** show the case  $g(\cdot)$  decreasing!

# Change of units

CHANGE-OF-UNITS TRANSFORMATION. Let  $X$  be a continuous random variable with distribution function  $F_X$  and probability density function  $f_X$ . If we change units to  $Y = rX + s$  for real numbers  $r > 0$  and  $s$ , then

$$F_Y(y) = F_X\left(\frac{y-s}{r}\right) \quad \text{and} \quad f_Y(y) = \frac{1}{r}f_X\left(\frac{y-s}{r}\right).$$

- For  $X \sim N(\mu, \sigma^2)$ , how is  $Z = \frac{X}{\sigma} + \frac{-\mu}{\sigma} = \frac{X-\mu}{\sigma}$  distributed?
- $f_Z(z) = \sigma f_X(\sigma z + \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$
- Hence,  $Z \sim N(0, 1)$
- In particular, for  $X \sim N(\mu, \sigma^2)$ , we have:

$$P(X \leq a) = P\left(Z \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

# Example: $\Lambda(\mu, \sigma^2)$

**Log-normal distribution**  $Y = e^X$  for  $X \sim N(\mu, \sigma^2)$ , i.e.,  $\log(Y) \sim N(\mu, \sigma^2)$

- $Y = g(X) = e^X$  Support is  $]0, \infty[$
- $g(x) = e^x$  is increasing, and  $g^{-1}(y) = \log y$ , and  $\frac{dg^{-1}(y)}{dy} = \frac{1}{y}$

$$F_Y(y) = F_X(g^{-1}(y)) = \Phi\left(\frac{\log y - \mu}{\sigma}\right) \quad f_Y(y) = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}$$

- $E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx = \int_{-\infty}^{\infty} yf_Y(y)dy = e^{\mu + \sigma^2/2}$
- Plausible and empirically adequate model for:
  - ▶ length of comments in posts, dwell time reading online articles, length of chess games, ...
  - ▶ size of living tissue, number of hospitalized cases in epidemics, blood pressure, ...
  - ▶ income of 97%–99% of the population, the number of citations, log of city size, ...
  - ▶ times to repair a maintainable system, size of audio-video files, amount of internet traffic per unit time, ...

**See R script**





# Example

- $X \sim U(0, 1)$  radius  $f_X(x) = 1$   $F_X(x) = x$  for  $x \in [0, 1]$

- $Y = g(X) = \pi \cdot X^2$

Support is  $[0, \pi]$

- $g(x) = \pi x^2$  is increasing, and  $g^{-1}(y) = \sqrt{\frac{y}{\pi}}$ , and  $\frac{dg^{-1}(y)}{dy} = \frac{1}{2\sqrt{\pi y}}$

$$F_Y(y) = F_X(g^{-1}(y)) = \sqrt{\frac{y}{\pi}} \quad f_Y(y) = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} = \frac{1}{2\sqrt{\pi y}}$$

*Do not lift distributions from a data column  
to a derived column in your data analytics project!*

**See R script**

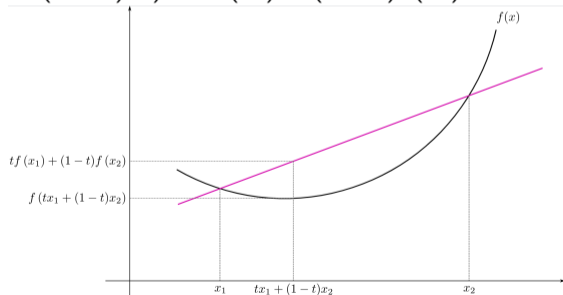
- Notice that:  $g(E[X]) = \pi/4 \leq E[g(X)] = \int_0^1 g(x)f_X(x)dx = \int_0^\pi yf_Y(y)dy = \frac{\pi}{3}$

# Jensen's inequality

JENSEN'S INEQUALITY. Let  $g$  be a convex function, and let  $X$  be a random variable. Then

$$g(E[X]) \leq E[g(X)].$$

- $f(\cdot)$  is convex if  $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$  for  $t \in [0, 1]$



- if  $f''(x) \geq 0$  then  $f(\cdot)$  is convex, e.g.,  $g(x) = \pi x^2$  or  $g(x) = 1/x$  for  $x \geq 0$

# Corollary and Example

**Corollary** (see [T, Ex. 8.11]). For a concave function  $g$ , namely  $g''(x) \leq 0$ :  $g(E[X]) \geq E[g(X)]$

- $\log(x)$  is concave since  $\log''(x) = -1/x^2 \leq 0$
- Let  $X$  be discrete with finite domain of  $n$  elements
  - ▶ By corollary above:

$$H(X) = E\left[\log \frac{1}{p(X)}\right] \leq \log E\left[\frac{1}{p(X)}\right]$$

- ▶ By change of variable:

$$E\left[\frac{1}{p(X)}\right] = \sum_i \frac{p(a_i)}{p(a_i)} = n$$

and then maximum entropy is:

$$H(X) \leq \log n$$

- ▶ E.g.,  $X \sim \text{Ber}(p)$ , maximum entropy (uncertainty) for equiprobable events  $p = 1/2$

# Variance

- **Investment A.**  $P(X = 450) = 0.5$     $P(X = 550) = 0.5$     $E[X] = 500$
- **Investment B.**  $P(X = 0) = 0.5$     $P(X = 1000) = 0.5$     $E[X] = 500$
- Spread around the mean is important!

## Variance and standard deviations

The *variance*  $Var(X)$  of a random variable  $X$  is the number:

$$Var(X) = E[(X - E[X])^2]$$

$\sigma_X = \sqrt{Var(X)}$  is called the *standard deviation* of  $X$ .

- The standard deviation has the same dimension as  $E[X]$  (and as  $X$ )
- For  $X$  discrete,  $Var(X) = \sum_i (a_i - E[X])^2 p(a_i)$
- **Investment A.**  $Var(X) = 50^2$  and  $\sigma_X = 50$
- **Investment B.**  $Var(X) = 500^2$  and  $\sigma_X = 500$

# Examples

- For  $a \in \mathbb{R}$ :

$$E[|X - a|] \leq \sqrt{E[(X - a)^2]}$$

- ▶ Apply Jensen's ineq. for  $g(y) = y^2$  convex on the r.v.  $Y = |X - a|$

- Median minimizes absolute deviation, i.e., for  $a \in \mathbb{R}$ :

$$E[|X - m_X|] \leq E[|X - a|]$$

- ▶ **Prove it!** (for continuous functions) Hint:  $\frac{d}{dx}|x| = x/|x|$

- Maximum distance between expectation and median:

$$|E[X] - m_X| \leq E[|X - m_X|] \leq E[|X - E[X]|] \leq \sqrt{E[(X - E[X])^2]} = \sigma_X$$

- ▶ Apply Jensen's ineq. for  $g(y) = |y|$  convex on the r.v.  $Y = X - m_X$  plus two results above

# Mode

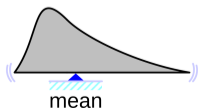
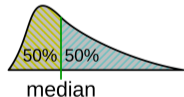
- For discrete r.v.  $X$  with p.m.f.  $p(\cdot)$ : the values  $a$  such that  $p(a)$  is maximum, i.e.:

$$\arg \max_a p(a)$$

- ▶ Can be more than one, e.g., in  $Ber(0.5)$
- For continuous r.v.  $X$  with d.f.  $f(\cdot)$ : the values  $x$  such that  $f(x)$  is a local maximum, e.g.:

$$f'(x) = 0 \quad \text{and} \quad f''(x) < 0$$

- ▶ Notice: **local** maximum!
- Unimodal distribution = that have only one mode



# Variance

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

**Proof.**

$$\begin{aligned}\text{Var}(X) &= E[(X - E[X])(X - E[X])] \\ &= E[X^2 + E[X]^2 - 2XE[X]] \\ &= E[X^2] + E[X]^2 - E[2XE[X]] \\ &= E[X^2] + E[X]^2 - 2E[X]E[X] = E[X^2] - E[X]^2\end{aligned}$$

- $E[X^2]$  is called the *second moment* of  $X$

$$\int_{-\infty}^{\infty} x^2 f(x) dx$$

**Corollary.**

$$\text{Var}(rX + s) = r^2 \text{Var}(X)$$

**Prove it!**

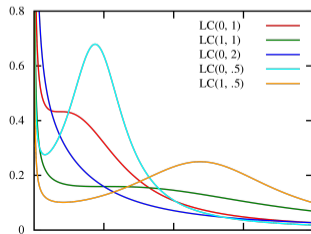
- Variance insensitive to shift  $s$ !

# Variance may be infinite or may not exist!

Standard deviation  $\sigma_X$  is a measure of the margin of error around a predicted value (e.g., temperature “ $20 \pm 1.5$ ”).

An infinite or non-existent margin of error is no prediction at all.

- Variance may not exist!
  - ▶ If expectation does not exist!
  - ▶ Also in cases when expectation exists
    - We'll see later *Power laws*.
- Variance can be infinite
  - ▶ Distributions have fat upper tails that decrease at an extremely slow rate.
  - ▶ The slow decay of probability increases the odds of very extreme values (*outliers*)
  - ▶ E.g.,  $e^X$  for  $X \sim \text{Cau}(0, 1)$



[log-Cauchy distribution]



# Variance

- Variance of some discrete distributions

- ▶  $X \sim U(m, M)$     $E[X] = \frac{(m+M)}{2}$     $Var(X) = \frac{(M-m+1)^2-1}{12}$ 
  - use  $Var(X) = Var(X - m)$ , call  $n = M - m + 1$  and  $\sum_{i=1}^{n-1} i^2 = \frac{(n-1)n(2n-1)}{6}$
- ▶  $X \sim Ber(p)$     $E[X] = p$     $Var(X) = p^2(1-p) + (1-p)^2p = p(1-p)$
- ▶  $X \sim Bin(n, p)$     $E[X] = n \cdot p$     $Var(X) = np(1-p)$ 
  - Because ... we'll see later
- ▶  $X \sim Geo(p)$     $E[X] = \frac{1}{p}$     $Var(X) = \frac{1-p}{p^2}$ 
  - Hint: use  $Var(X) = E[X^2] - E[X]^2$  and  $\sum_{k=1}^{\infty} k^2 \cdot x^{k-1} = \frac{1+x}{(1-x)^3}$
- ▶  $X \sim NBin(n, p)$     $E[X] = \frac{n \cdot p}{1-p}$     $Var(X) = n \frac{1-p}{p^2}$ 
  - Because ... we'll see later
- ▶  $X \sim Poi(\mu)$     $E[X] = \mu$     $Var(X) = \mu$ 
  - Because, when  $n \rightarrow \infty$ :  $Bin(n, \mu/n) \rightarrow Poi(\mu)$

See [seeing-theory.brown.edu](http://seeing-theory.brown.edu)

# Variance

- Variance of some continuous distributions
  - ▶  $X \sim U(\alpha, \beta)$     $E[X] = (\alpha + \beta)/2$     $\text{Var}(X) = (\beta - \alpha)^2/12$ 
    - **Prove it!** Recall that  $f(x) = 1/(\beta - \alpha)$
  - ▶  $X \sim \text{Exp}(\lambda)$     $E[X] = 1/\lambda$     $\text{Var}(X) = 1/\lambda^2$ 
    - **Prove it!** Recall that  $f(x) = \lambda e^{-\lambda x}$
  - ▶  $X \sim N(\mu, \sigma^2)$     $E[X] = \mu$     $\text{Var}(X) = \sigma^2$ 
    - **Prove it!** Hint: use  $z = \frac{x - \mu}{\sigma}$  and integration by parts.
  - ▶  $X \sim \text{Erl}(n, \lambda)$     $E[X] = n/\lambda$     $\text{Var}(X) = n/\lambda^2$ 
    - Because ... we'll see later

# $E[\cdot]$ and $Var(\cdot)$ of random variables with bounded support

Assume  $a \leq X \leq b$ , or more generally  $P(a \leq X \leq b) = 1$

[almost surely or a.s.]

It turns out that expectation and variance are finite!

- $a \leq E[X] \leq b$ 
  - ▶ E.g., for  $X$  continuous,  $E[X] = \int_a^b xf(x)dx \leq \int_a^b bf(x)dx = b$
- $0 \leq Var(X) \leq (b-a)^2/4$

## Proof.

- ▶ For  $\gamma \in \mathbb{R}$ , consider  $E[(X - \gamma)^2] = \gamma^2 - 2\gamma E[X] + E[X^2]$
- ▶ It is minimum for  $\gamma = E[X]$  (consider  $\frac{d}{d\gamma}(\gamma^2 - 2\gamma a + b)$ )
- ▶ Thus,  $E[(X - E[X])^2] = Var(X) \leq E[(X - \gamma)^2]$
- ▶ Since  $(X - \gamma)^2 \leq (b - \gamma)^2$ , for  $\gamma = (a+b)/2$  we have:

$$Var(X) \leq E[(X - \frac{(a+b)}{2})^2] \leq (b - \frac{(a+b)}{2})^2 = \frac{(b-a)^2}{4}$$

- **Exercise at home:** show that the bound  $(b-a)^2/4$  can be reached.