Statistical Methods for Data Science
Lesson 06 - Expectation and variance. Computations with random variables.

Salvatore Ruggieri

Department of Computer Science
University of Pisa
salvatore.ruggieri@unipi.it
Expectation of a discrete random variable

• Buy lottery ticket every week, \( p = 1/10000 \)

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

• 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 \( a_1, a_2, \ldots \) 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

Look at [seeing-theory.brown.edu](http://seeing-theory.brown.edu)
Expected value may be infinite!

- $X$ with PMF $p(2^k) = 2^{-k}$ for $k = 1, 2, \ldots$
- $p()$ is a PMF 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 of some discrete distributions
  - $X \sim U(m, M)$  $E[X] = (m+M)/2$
    - $\sum_{i=m}^{M} \frac{i}{M-m+1} = \ldots$
  - $X \sim Ber(p)$  $E[X] = p$
  - $X \sim Bin(n, p)$  $E[X] = n \cdot p$
    - Because . . . we’ll see later
  - $X \sim NBin(n, p)$  $E[X] = \frac{n \cdot p}{1-p}$
    - Because . . . we’ll see later
  - $X \sim Poi(\mu)$  $E[X] = \mu$
    - Because, when $n \to \infty$: $Bin(n, \mu/n) \to Poi(\mu)$
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**
  - $X \sim U(\alpha, \beta) \quad E[X] = (\alpha + \beta)/2$
  - $X \sim \text{Exp}(\lambda) \quad E[X] = 1/\lambda$
    - Because $\int_{0}^{\infty} x \lambda e^{-\lambda x} \, dx = [-e^{-\lambda x}(x + 1/\lambda)]_{0}^{\infty} = e^{0}(0 + 1/\lambda)$
  - $X \sim N(\mu, \sigma^2) \quad E[X] = \mu$
    - Because: $\int_{-\infty}^{\infty} x \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \, dx = \mu + \int_{-\infty}^{\infty} (x - \mu) \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \, dx = \mu + \sigma \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \, dz = \mu$
Expected value may not exists!

- Cauchy distribution
  \[ f(x) = \frac{1}{\pi(1 + x^2)} \]

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

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

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

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

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

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

See R script
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?

\[
\frac{2}{\text{sum}(v \times p)}
\]
\[
\text{sum}(\frac{2}{v} \times p)
\]

\[ X = \text{velocity}, \quad g(X) = 2/X \quad \text{time to reach school} \]
\[ E[g(X)] \neq g(E[X]) \]
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, $f_g(a) = dF_g(a)/da = 1/20 \sqrt{a}$
- $E[g(X)] = \frac{1}{20} \int_0^{100} \frac{x}{\sqrt{x}} dx = \frac{1}{20} \frac{2}{3} \left[ x^{3/2} \right]_0^{100} = 100/3$
- Alternatively, $E[g(X)] = \int_0^{10} x^2 \frac{1}{10} dx = \frac{1}{10} \frac{1}{3} \left[ x^3 \right]_0^{10} = 100/3$

See R script.
Theorem (Change of units)

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

- Prove it!
- Corollary:

\[ E[X - E[X]] = E[X] - E[X] = 0 \]
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

$$P_Y(Y = y) = \begin{cases} 
150/200 & \text{if } y = 0 \\
1/200 & \text{if } 1 \leq y \leq 50 
\end{cases}$$

$g^{-1}(0) = \{1, \ldots, 150\}$

$g^{-1}(y) = \{y + 150\}$

- 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 random variables

**Theorem**

For a continuous random variable $X$, the density functions of $Y = g(X)$ when $g()$ 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 $g()$ increasing) Since $g()$ 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:

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

Show the case $g()$ decreasing!
Example

- $X \sim U(0,1)$ radius \quad f_X(x) = 1 \quad F_X(x) = x$ for $x \in [0,1]$
- $Y = g(X) = \pi \cdot X^2$
- $g(x) = \pi x^2$ 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

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

- $f()$ 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 $f()$ is convex, e.g., $g(x) = \pi x^2$ or $g(x) = \frac{1}{x}$ for $x \geq 0$
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?
- $f_Z(z) = \sigma f_X(\sigma y + \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$
- Hence, $Z \sim N(0, 1)$
- In particular, any probability for $X$ can be expressed in terms of probability for $Z$:

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

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

### Variance and standard deviations

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

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

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

- The standard deviation has the same dimension as $E[X]$ (and as $X$)
- **Investment A.** $\text{Var}(X) = 50^2$ and $\sigma_X = 50$
- **Investment B.** $\text{Var}(X) = 500^2$ and $\sigma_X = 500$
Variance

• It holds that:

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

• \( E[X^2] \) is called the second moment of \( X \)

• Prove it!

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

• Corollary:

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

• Prove it!

• Variance insensitive to shift \( s \)!
• Variance may not exists!
  ▶ If expectation does exists!
  ▶ Also in cases when expectation exists
    □ We’ll see later
• Variance of some discrete distributions
  ▶ $X \sim U(m, M)$  $E[X] = \frac{(m+M)}{2}$  $\text{Var}(X) = \frac{(M-m+1)^2-1}{12}$
    □ use $\text{Var}(X) = \text{Var}(X - m)$, call $n = M - m + 1$ and $\sum_{i=1}^{n-1} i^2 = \frac{(n-1)n(2n-1)}{6}$
  ▶ $X \sim \text{Ber}(p)$  $E[X] = p$  $\text{Var}(X) = p^2(1-p) + (1-p)^2 p = p(1-p)$
  ▶ $X \sim \text{Bin}(n, p)$  $E[X] = n \cdot p$  $\text{Var}(X) = np(1-p)$
    □ Because . . . we’ll see later
  ▶ $X \sim \text{Geo}(p)$  $E[X] = \frac{1}{p}$  $\text{Var}(X) = \frac{1-p}{p^2}$
    □ Hint: use $\text{Var}(X) = E[X^2] - E[X]^2$ and $\sum_{k=1}^{\infty} k^2 \cdot x^{k-1} = \frac{1+x}{(1-x)^3}$
  ▶ $X \sim \text{NBin}(n, p)$  $E[X] = \frac{n \cdot p}{1-p}$  $\text{Var}(X) = n \frac{1-p}{p^2}$
    □ Because . . . we’ll see later
  ▶ $X \sim \text{Poi}(\mu)$  $E[X] = \mu$  $\text{Var}(X) = \mu$
    □ Because, when $n \to \infty$: $\text{Bin}(n, \mu/n) \to \text{Poi}(\mu)$

Look at [seeing-theory.brown.edu](http://seeing-theory.brown.edu)
• Variance of some continuous distributions
  ▸ $X \sim U(\alpha, \beta)$ \quad $E[X] = (\alpha + \beta)/2$ \quad $Var(X) = (\beta - \alpha)^2/12$
    \begin{itemize}
      \item Prove it! Recall that $f(x) = 1/(\beta - \alpha)$
    \end{itemize}
  ▸ $X \sim Exp(\lambda)$ \quad $E[X] = 1/\lambda$ \quad $Var(X) = 1/\lambda^2$
    \begin{itemize}
      \item Prove it! Recall that $f(x) = \lambda e^{-\lambda x}$
    \end{itemize}
  ▸ $X \sim N(\mu, \sigma^2)$ \quad $E[X] = \mu$ \quad $Var(X) = \sigma^2$
    \begin{itemize}
      \item Prove it! Hint: use $z = \frac{x-\mu}{\sigma}$ and integration by parts.
    \end{itemize}