Master Program in *Data Science and Business Informatics*

Statistics for Data Science

Lesson 10 - Moments. Functions of random variables

Salvatore Ruggieri

Department of Computer Science
University of Pisa, Italy
salvatore.ruggieri@unipi.it

Moments

- Let X be a continuous random variable with density function $f(x)$
- k^{th} *moment* of X, if it exists, is:

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

- $\mu = E[X]$ is the first moment of X
- k^{th} *central moment* of X is:

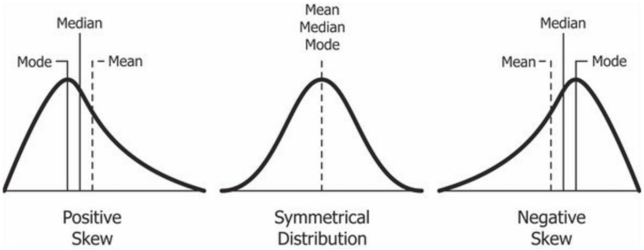
$$\mu_k = E[(X-\mu)^k] = \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx$$

- $\sigma = \sqrt{E[(X-\mu)^2]}$ standard deviation is the square root of the second central moment
- k^{th} *standardized moment* of X is:

$$\tilde{\mu}_k = \frac{\mu_k}{\sigma^k} = E\left[(\frac{X-\mu}{\sigma})^k\right]$$

Skewness

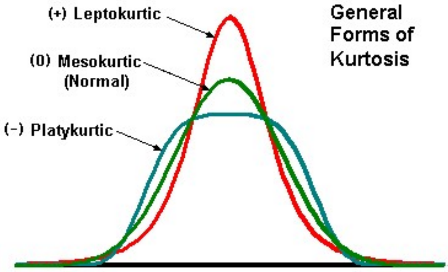
- $\tilde{\mu}_1 = {}^{E[(X-\mu)]}/_{\sigma} = 0$ since $E[X-\mu] = 0$
- $\tilde{\mu}_2 = {}^{E[(X-\mu)^2]}/_{\sigma^2} = 1$ since $\sigma^2 = E[(X-\mu)^2]$
- $\tilde{\mu}_3 = {}^{E[(X-\mu)^3]}/_{\sigma^3}$ *[(Pearson's moment) coefficient of skewness]*
- Skewness indicates direction and magnitude of a distribution's deviation from symmetry



- E.g., for $X \sim Exp(\lambda)$, $\tilde{\mu}_3 = 2$ **Prove it!**

Kurtosis

- $\tilde{\mu}_4 = E[(\frac{X-\mu}{\sigma})^4]$ *[(Pearson's moment) coefficient of kurtosis]*
- For $X \sim N(\mu, \sigma)$, $\tilde{\mu}_4 = 3$ $\tilde{\mu}_4 - 3$ *is called kurtosis in excess*
- Kurtosis is a measure of the dispersion of X around the two values $\mu \pm \sigma$



- $\tilde{\mu}_4 > 3$ *Leptokurtic* (slender) distribution has *fatter* tails. May have outlier problems.
- $\tilde{\mu}_4 < 3$ *Platykurtic* (broad) distribution has *thinner* tails

See R script

Functions of two or more random variables: expectation

- $V = \pi H R^2$ be the volume of a vase of height H and radius R
- $g(H, R) = \pi H R^2$ is a random variable (function of random variables)
- $P_V(V = 3) = P_{HR}(\pi H R^2 = 3)$
- How to calculate $E[V]$?

TWO-DIMENSIONAL CHANGE-OF-VARIABLE FORMULA. Let X and Y be random variables, and let $g : \mathbb{R}^2 \to \mathbb{R}$ be a function.
If X and Y are *discrete* random variables with values $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$, respectively, then

$$\mathrm{E}[g(X,Y)] = \sum_i \sum_j g(a_i, b_j) \mathrm{P}(X = a_i, Y = b_j).$$

If X and Y are *continuous* random variables with joint probability density function f, then

$$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) \,\mathrm{d}x \,\mathrm{d}y.$$

If $H \perp\!\!\!\perp R$:

$$E[V] = E[\pi H R^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \pi h r^2 f_H(h) f_R(r) dh dr$$

Linearity of expectations

Theorem. For X and Y random variables, and $s, t \in \mathbb{R}$:

$$E[rX + sY + t] = rE[X] + sE[Y] + t$$

Proof. (discrete case)

$$E[rX + Ys + t] = \sum_a \sum_b (ra + sb + t) P(X = a, Y = b)$$

$$= \left(r \sum_a \sum_b aP(X = a, Y = b) \right) + \left(s \sum_a \sum_b bP(X = a, Y = b) \right) + \left(t \sum_a \sum_b P(X = a, Y = b) \right)$$

$$= \left(r \sum_a aP(X = a) \right) + \left(s \sum_b bP(Y = b) \right) + t = rE[X] + sE[Y] + t$$

Corollary. $E[a_0 + \sum_{i=1}^n a_i X_i] = a_o + \sum_{i=1}^n a_i E[X_i]$

Corollary. $X \leq Y$ implies $E[X] \leq E[Y]$
Proof. $Z = Y - X \geq 0$ implies $E[Z] = E[Y] - E[X] \geq 0$, i.e., $E[Y] \geq E[X]$.

Applications

- Expectation of some discrete distributions
  - $X \sim Ber(p) \quad E[X] = p$
  - $X \sim Bin(n, p) \quad E[X] = n \cdot p$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Ber(p)$
  - $X \sim Geo(p) \quad E[X] = \frac{1}{p}$
  - $X \sim NBin(n, p) \quad E[X] = \frac{n \cdot (1-p)}{p}$
    - Because $X = \sum_{i=1}^{n} X_i - n$ for $X_1, \ldots, X_n \sim Geo(p)$
- Expectation of some continuous distributions
  - $X \sim Exp(\lambda) \quad E[X] = 1/\lambda$
  - $X \sim Erl(n, \lambda) \quad E[X] = \frac{n}{\lambda}$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Exp(\lambda)$

Expectation of product and quotients

Theorem. For $X \perp\!\!\!\perp Y$, we have: $E[XY] = E[X]E[Y]$ **Prove it!**

PROPAGATION OF INDEPENDENCE. Let $X_1, X_2, \ldots, X_n$ be independent random variables. For each i, let $h_i : \mathbb{R} \to \mathbb{R}$ be a function and define the random variable

$$Y_i = h_i(X_i).$$

Then $Y_1, Y_2, \ldots, Y_n$ are also independent.

Corollary. For $X \perp\!\!\!\perp Y$ and $Y > 0$, we have: $E[X/Y] \geq E[X]/E[Y]$
Proof. $X \perp\!\!\!\perp Y$ implies $X \perp\!\!\!\perp {}^1\!/_Y$. By theorem above:

$$E[X/Y] = E[X \cdot {}^1\!/_Y] = E[X]E[{}^1\!/_Y] \geq E[X]/E[Y]$$

because by Jensen's inequality $E[{}^1\!/_Y] \geq 1/E[Y]$ since $1/y$ is convex for y 0. $\square$

Exercise at home. Show that $E[X/Y] = E[X]/E[Y]$ is a false claim.

Law of iterated/total expectation

Conditional expectation

$$E[X|Y=b] = \sum_i a_i p(a_i|b) \qquad E[X|Y=y] = \int_{-\infty}^{\infty} x f(x|y) dx$$

Theorem. (Law of iterated/total expectation)

$$E_Y[E[X|Y]] = E[X]$$

Proof. (for X, Y discrete random variables)

$$E_Y[E[X|Y]] = \sum_j \sum_i a_i p_{X|Y}(a_i|b_j) p_Y(b_j) = \sum_j \sum_i a_i p_{XY}(a_i, b_j) = \sum_i a_i p_X(a_i) = E[X]$$

Example (cfr the example from Lesson 1 on the Law of total probability)

- Factory 1's light bulbs working hours $\sim Exp(1/1000)$
- Factory 2's light bulbs working hours $\sim Exp(1/2000)$
- Factory 1 supplies 60% of the total bulbs on the market and Factory 2 supplies 40% of it.
- *What is the average work hour of a light bulb on the market?*

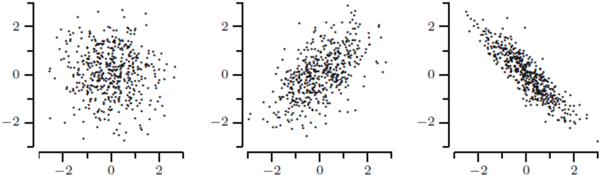
Variance of the sum and covariance

$$
\begin{aligned}
Var(X+Y) &= E[(X+Y-E[X+Y])^2] = E[((X-E[X])+(Y-E[Y]))^2] \\
&= E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2\color{red}{E[(X-E[X])(Y-E[Y])]} \\
&= Var(X) + Var(Y) + 2Cov(X,Y)
\end{aligned}
$$

Covariance

The *covariance* $Cov(X,Y)$ of two random variables X and Y is the number:

$$Cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$



Uncorrelated **Positively correlated** **Negatively correlated**

Covariance

Theorem. $Cov(X, Y) = E[XY] - E[X]E[Y]$ **Prove it!**

- If X and Y are independent ($X \perp\!\!\!\perp Y$):

$$Cov(X, Y) = 0 \qquad Var(X + Y) = Var(X) + Var(Y)$$

- But there are X and Y uncorrelated (ie., $Cov(X, Y) = 0$) that are dependent!
- Variances of some discrete distributions
  - $X \sim Ber(p)$ $\quad Var(X) = p(1 - p)$
  - $X \sim Bin(n, p)$ $\quad Var(X) = np(1 - p)$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Ber(p)$ and independent
  - $X \sim Geo(p)$ $\quad Var(X) = \frac{1-p}{p^2}$
  - $X \sim NBin(n, p)$ $\quad Var(X) = n\frac{1-p}{p^2}$
    - Because $X = \sum_{i=1}^{n} X_i - n$ for $X_1, \ldots, X_n \sim Geo(p)$ and independent
- Variances of some continuous distributions
  - $X \sim Exp(\lambda)$ $\quad Var(X) = 1/\lambda^2$
  - $X \sim Erl(n, \lambda)$ $\quad Var(X) = \frac{n}{\lambda^2}$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Exp(\lambda)$ and independent

Covariance and covariance matrix

COVARIANCE UNDER CHANGE OF UNITS. Let X and Y be two random variables. Then

$$\mathrm{Cov}(rX + s, tY + u) = rt\,\mathrm{Cov}(X, Y)$$

for all numbers r, s, t, and u.

- Hence, $Var(rX + sY + t) = r^2 Var(X) + s^2 Var(Y) + 2rsCov(X, Y)$
- **Bivariate** Normal/Gaussian distribution:

$$(X, Y) \sim N((\mu_X, \mu_X), \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix})$$

  - where marginals are $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, and $Cov(X, Y) = \sigma_{XY}$
  - Covariance matrix $\mathbf{\Sigma}_{ij} = Cov(X_i, X_j)$ for a vector $\mathbf{X} = (X_1, \ldots, X_n)$ of r.v.'s
- Covariance depends on the unit of measure!

See R script lesson 08

Correlation coefficient

DEFINITION. Let X and Y be two random variables. The *correlation coefficient* $\rho(X, Y)$ is defined to be 0 if $\mathrm{Var}(X) = 0$ or $\mathrm{Var}(Y) = 0$, and otherwise

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

- Correlation coefficient is *dimensionless* (not affected by change of units)
  - E.g., if X and Y are in Km, then $Cov(X, Y)$, $Var(X)$ and $Var(Y)$ are in Km^2
- Moreover: $-1 \leq \rho(X, Y) \leq 1$
  - The bounds are derived from the **Cauchy–Schwarz's inequality**:

$$E[|XY|] \leq \sqrt{E[X^2]}\sqrt{E[Y^2]}$$

Proof. For any $u, w \in \mathbb{R}$, we have $2|uw| \leq u^2 + w^2$. Therefore, $2|UW| \leq U^2 + W^2$ for r.v.'s U and V. By defining $U = {}^{X}/_{\sqrt{E[X^2]}}$ and $W = {}^{Y}/_{\sqrt{E[Y^2]}}$ $^{(*)}$, we have $2 \cdot {}^{|XY|}/_{\sqrt{E[X^2]}\sqrt{E[Y^2]}} \leq {}^{X^2}/_{E[X^2]} + {}^{Y^2}/_{E[Y^2]}$. Taking the expectations, we conclude: $2 \cdot {}^{E[|XY|]}/_{\sqrt{E[X^2]}\sqrt{E[Y^2]}} \leq 2$. $\square$

(*) The case $E[X^2] = 0$ or $E[Y^2] = 0$ is left as an exercise.

Bivariate Normal/Gaussian distribution

$$(X, Y) \sim N((\mu_X, \mu_Y), \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix})$$

where marginals are $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, and $Cov(X, Y) = \sigma_{XY}$

- Since $\sigma_{XY} = \rho(X, Y) \cdot \sigma_X \cdot \sigma_Y$:

$$(X, Y) \sim N((\mu_X, \mu_Y), \begin{pmatrix} \sigma_X^2 & \rho(X, Y) \cdot \sigma_X \cdot \sigma_Y \\ \rho(X, Y) \cdot \sigma_X \cdot \sigma_Y & \sigma_Y^2 \end{pmatrix})$$

- Density of $N((0, 0), (1, \sigma_{XY}, \sigma_{XY}, 1))$:

$$f(x, y) = \frac{1}{2\pi\sqrt{1 - \sigma_{XY}^2}} e^{-\frac{1}{2(1-\sigma_{XY}^2)}(x^2+y^2-2xy\sigma_{XY})}$$

- Useful facts for (X, Y) bivariate Normal:
  - for (X, Y) bivariate Normal: $\rho(X, Y) = 0$ iff $X \perp\!\!\!\perp Y$, i.e., uncorrelation equals independence
  - (X, Y) bivariate Normal iff $aX + bY$ is Normal for any $a, b \in \mathbb{R}$

Sum of independent Normal random variables

- See Lesson 04 and Lesson 08 for convolution formulas

ADDING TWO INDEPENDENT CONTINUOUS RANDOM VARIABLES. Let X and Y be two independent continuous random variables, with probability density functions f_X and f_Y. Then the probability density function f_Z of $Z = X + Y$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\,\mathrm{d}y$$

for $-\infty < z < \infty$.

Theorem. If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ and $X \perp\!\!\!\perp Y$, then:

$$Z = X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Proof. See [T, Sect. 11.2] $\square$

- In general: $Z = rX + sY + t \sim N(r\mu_X + s\mu_Y + t, r^2\sigma_X^2 + s^2\sigma_Y^2)$
- The converse of the theorem also holds: [**Lévy-Cramér theorem**]
  - If $X \perp\!\!\!\perp Y$ and $Z = X + Y$ is normally distributed, then X and Y follow a normal distribution.

Extremes of independent random variables

THE DISTRIBUTION OF THE MAXIMUM. Let $X_1, X_2, \ldots, X_n$ be n independent random variables with the same distribution function F, and let $Z = \max\{X_1, X_2, \ldots, X_n\}$. Then

$$F_Z(a) = (F(a))^n.$$

- $P(Z \leq a) = P(X_1 \leq a, \ldots, X_n \leq a) = \prod_{i=1}^{n} P(X_i \leq a) = ((F(a))^n$
- Example: maximum water level over 365 days assuming water level on a day is $U(0,1)$
- Example: maximum of two rolls of **a die with 4 sides**

THE DISTRIBUTION OF THE MINIMUM. Let $X_1, X_2, \ldots, X_n$ be n independent random variables with the same distribution function F, and let $V = \min\{X_1, X_2, \ldots, X_n\}$. Then

$$F_V(a) = 1 - (1 - F(a))^n.$$

- $P(V \leq a) = 1 - P(X_1 > a, \ldots, X_n > a) = 1 - \prod_{i=1}^{n}(1 - P(X_i \leq a) = 1 - ((1 - F(a))^n$

Product and quotient of independent random variables

PRODUCT OF INDEPENDENT CONTINUOUS RANDOM VARIABLES. Let X and Y be two independent continuous random variables with probability densities f_X and f_Y. Then the probability density function f_Z of $Z = XY$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y\left(\frac{z}{x}\right) f_X(x) \frac{1}{|x|} \, \mathrm{d}x$$

for $-\infty < z < \infty$.

QUOTIENT OF INDEPENDENT CONTINUOUS RANDOM VARIABLES. Let X and Y be two independent continuous random variables with probability densities f_X and f_Y. Then the probability density function f_Z of $Z = X/Y$ is given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(zx) f_Y(x) |x| \, \mathrm{d}x$$

for $-\infty < z < \infty$.

- $X, Y \sim N(0,1)$ independent, $Z = X/Y \sim Cau(0,1)$ where:

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