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:

$$E[X^k] = \int_{1}^{1} x^k f(x)dx$$

$= E[X]$ is the first moment of X

k^{th} *central moment* of X is:

$$_k = E[(X \quad)^k] = \int_{1}^{1} (x \quad)^k f(x)dx$$

$= \sqrt{E[(X \quad)^2]}$ standard deviation is the square root of the second central moment

k^{th} *standardized moment* of X is:

$$\tilde{}_k = \frac{k}{k} = E\left(\frac{X}{\quad}\right)^k$$

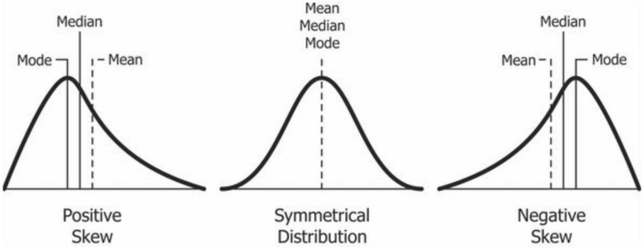
Skewness

$\tilde{\ }_1 = E[(X \quad)] = \quad = 0$ since $E[X \quad] = 0$

$\tilde{\ }_2 = E[(X \quad)^2] = {}^2 = 1$ since ${}^2 = E[(X \quad)^2]$

$\tilde{\ }_3 = E[(X \quad)^3] = {}^3$ *[(Pearson's moment) coefficient of skewness]*

Skewness indicates direction and magnitude of a distribution's deviation from symmetry



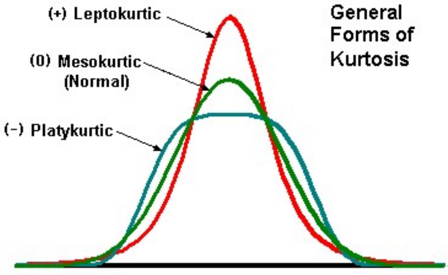
E.g., for $X \quad Exp(\ )$, $\tilde{\ }_3 = 2$ **Prove it!**

Kurtosis

$\tilde{}_4 = E[(\frac{X \quad}{\quad})^4]$ *[(Pearson's moment) coefficient of kurtosis]*

For $X \quad N(\ ; \ )$, $\tilde{}_4 = 3$ $\tilde{}_4 \quad 3$ *is called kurtosis in excess*

Kurtosis is a measure of the dispersion of X around the two values



$\tilde{}_4 > 3$ *Leptokurtic* (slender) distribution has *fatter* tails. May have outlier problems.

$\tilde{}_4 < 3$ *Platykurtic* (broad) distribution has *thinner* tails

See R script

Functions of two or more random variables: expectation

$V = \quad HR^2$ be the volume of a vase of height H and radius R
$g(H; R) = \quad HR^2$ is a random variable (function of random variables)
$P_V(V = 3) = P_{HR}(\quad HR^2 = 3)$
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 ?? R:

$$E[V] = E[\quad HR^2] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \quad hr^2 f_H(h) f_R(r) dh dr$$

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)

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

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

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

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

Corollary. $X \geq Y$ implies $E[X] \geq E[Y]$
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

- $X \sim Ber(p) \Rightarrow E[X] = p$
- $X \sim Bin(n; p) \Rightarrow E[X] = n \cdot p$

  Because $X = \sum_{i=1}^{n} X_i$ for $X_1; \ldots; X_n \sim Ber(p)$
- $X \sim Geo(p) \Rightarrow E[X] = \frac{1}{p}$
- $X \sim NBin(n; p) \Rightarrow E[X] = \frac{n \cdot (1 - p)}{p}$

  Because $X = \sum_{i=1}^{n} X_i - n$ for $X_1; \ldots; X_n \sim Geo(p)$

Expectation of some continuous distributions

- $X \sim Exp(\lambda) \Rightarrow E[X] = 1/\lambda$
- $X \sim Erl(n; \lambda) \Rightarrow E[X] = \frac{n}{\lambda}$

  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 ?? 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 ?? Y and $Y > 0$, we have: $E[X{=}Y] \quad E[X]{=}E[Y]$
Proof. X ?? Y implies X ?? $1{=}Y$. By theorem above:

$$E[X{=}Y] = E[X \quad 1{=}Y] = E[X]E[1{=}Y] \quad E[X]{=}E[Y]$$

because by Jensen's inequality $E[1{=}Y] \quad 1{=}E[Y]$ since $1{=}y$ is convex for $y \, 0$.

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?

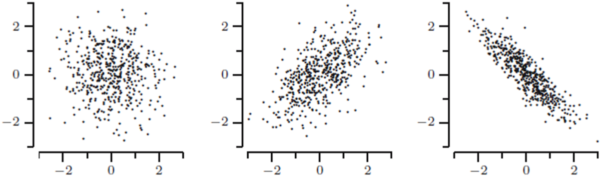
Variance of the sum and covariance

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

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. $Cov(X; Y) = E[XY] \quad E[X]E[Y]$ **Prove it!**

If X and Y are independent (X ⁇ 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

- $X \quad Ber(p) \quad Var(X) = p(1 \quad p)$
- $X \quad Bin(n; p) \quad Var(X) = np(1 \quad p)$
  Because $X = \sum_{i=1}^{n} X_i$ for $X_1; : : : ; X_n \quad Ber(p)$ and independent
- $X \quad Geo(p) \quad Var(X) = \frac{1 \quad p}{p^2}$
- $X \quad NBin(n; p) \quad Var(X) = n\frac{1 \quad p}{p^2}$
  Because $X = \sum_{i=1}^{n} X_i \quad n$ for $X_1; : : : ; X_n \quad Geo(p)$ and independent

Variances of some continuous distributions

- $X \quad Exp(\;) \quad Var(X) = 1 = \; ^2$
- $X \quad Erl(n; \;) \quad Var(X) = \frac{n}{\; ^2}$
  Because $X = \sum_{i=1}^{n} X_i$ for $X_1; : : : ; X_n \quad Exp(\;)$ 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:

$$(X;Y) \quad N((\ _X;\ \ _X);\quad \begin{matrix} {}^2_X & {}_{XY} \\ {}_{XY} & {}^2_Y \end{matrix}\quad)$$

- where marginals are $X \quad N(\ _X;\ {}^2_X)$, $Y \quad N(\ _Y;\ {}^2_Y)$, and $Cov(X;Y) = \ _{XY}$
- Covariance matrix $\mathbf{\Sigma}_{ij} = Cov(X_i;X_j)$ for a vector $\mathbf{X} = (X_1;:::;X_n)$ 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: $-1 \le \rho(X,Y) \le 1$

- The bounds are derived from the **Cauchy–Schwarz's inequality**:

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

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

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

Bivariate Normal/Gaussian distribution

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

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

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

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 $X \quad N(\ _X;\ ^2_X)$ and $Y \quad N(\ _Y;\ ^2_Y)$ and X ? Y, then:

$$Z = X + Y \quad N(\ _X +\ _Y;\ ^2_X +\ ^2_Y)$$

Proof. See [T, Sect. 11.2]

In general: $Z = rX + sY + t \quad N(r\ _X + s\ _Y + t; r^2\ ^2_X + s^2\ ^2_Y)$

The converse of the theorem also holds: **[Lévy-Cramér theorem]**

- If X ? 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

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

$$P(Z \le a) = P(X_1 \le a, \ldots, X_n \le a) = \prod_{i=1}^n P(X_i \le a) = ((F(a))^n$$

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

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

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