Statistical Methods for Data Science

Lesson 09 - Moments, joint distributions, sum of random variables

Salvatore Ruggieri

Department of Computer Science
University of Pisa
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:

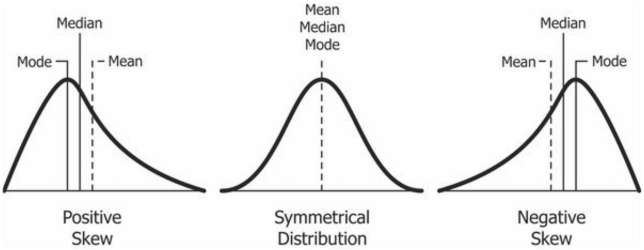
$$\mu_k = E[(X-\mu)^k] = \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx$$

- $\sigma = \sqrt{E[(X-\mu)^k]}$ 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

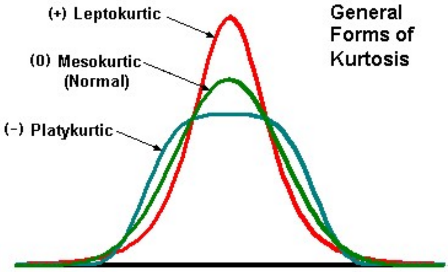
- $\tilde{\mu}_1 = {}^{E[(X-\mu)]}/_{\sigma} = 0$ since $E[X-\mu] = 0$
- $\tilde{\mu}_2 = {}^{E[(X-\mu)^2]}/_{\sigma^2} = 1$ since $\sigma^2 = E[(X-\mu)^2]$
- $\tilde{\mu}_3 = {}^{E[(X-\mu)^3]}/_{\sigma^3}$ *[(Pearson's moment) coefficient of skewness]*
- Skewness indicates direction and magnitude of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

Joint distributions

- Random variables related to the same experiment often influence one another
- $\Omega = \{(i, j) \mid i, j \in 1, \ldots, 6\}$ rolls of two dies
- $X = sum(i, j)$ and $Y = max(i, j)$
- $P(X = 4, Y = 3) = P(\{X = 4\} \cap \{Y = 3\}) = P(\{(3, 1), (1, 3)\}) = 2/36$

Joint and marginal p.m.f.

- In general:

$$P_{XY}(X = a, Y = b) = P(\{\omega \in \Omega \mid X(\omega) = a \text{ and } Y(\omega) = b\})$$

DEFINITION. The *joint probability mass function* p of two discrete random variables X and Y is the function $p : \mathbb{R}^2 \to [0,1]$, defined by

$$p(a,b) = \mathrm{P}(X = a, Y = b) \quad \text{for } -\infty < a, b < \infty.$$

- The marginal p.m.f.'s can be derived from the joint p.m.f. as:

$$p_X(a) = P_X(X = a) = \sum_b P_{XY}(X = a, Y = b)$$

$$p_Y(b) = P_Y(Y = b) = \sum_a P_{XY}(X = a, Y = b)$$

See R script

Joint and marginal CDF

- In general: $P_{XY}(X \leq a, Y \leq b) = P(\{\omega \in \Omega \,|X(\omega) \leq a \text{ and } Y(\omega) \leq b\})$

DEFINITION. The *joint distribution function* F of two random variables X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ defined by

$$F(a, b) = \mathrm{P}(X \leq a, Y \leq b) \quad \text{for } -\infty < a, b < \infty.$$

- The marginal distribution functions of X and Y are:

$$F_X(a) = P_X(X \leq a) = F_{XY}(a, \infty) = \lim_{b \to \infty} F(a, b)$$

$$F_Y(b) = P_Y(Y \leq b) = F_{XY}(\infty, b) = \lim_{a \to \infty} F(a, b)$$

- But given $F_X()$ and $F_Y()$ we cannot reconstruct $F_{XY}()$!

See R script

Joint distributions: continuous random variables

DEFINITION. Random variables X and Y have a *joint continuous distribution* if for some function $f : \mathbb{R}^2 \to \mathbb{R}$ and for all numbers a_1, a_2 and b_1, b_2 with $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$\mathrm{P}(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y)\, \mathrm{d}x\, \mathrm{d}y.$$

The function f has to satisfy $f(x, y) \geq 0$ for all x and y, and $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1$. We call f the *joint probability density function* of X and Y.

- The marginal density functions of X and Y are:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

- Moreover, as in the univariate case:

$$F(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x, y) dx dy \qquad f(x, y) = \frac{d}{dx}\frac{d}{dy} F(x, y) = \frac{d^2}{dxdy} F(x, y)$$

See R script

Independence of two random variables

- Conditional probability:

$$P(X \leq a | Y \leq b) = \frac{P(X \leq a, Y \leq b)}{P(Y \leq b)}$$

- Independence

$$P(X \leq a | Y \leq b) = P(X \leq a)$$

or, equivalently:

$$P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b) \qquad F_{XY}(a, b) = F_X(a) \cdot F_Y(b)$$

or, for discrete/continuous random variables, equivalently:

$$p(a, b) = p(a) \cdot p(b) \qquad f_{XY}(x, y) = f_X(x) \cdot f_Y(y)$$

Functions of random variables

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

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

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

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

Linearity of expectations

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

Proof. (discrete case)

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

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

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

□

- If X and Y are independent, $E[XY] = E[X]E[Y]$ **Prove it!**

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
  - $X \sim Exp(\lambda) \quad E[X] = 1/\lambda$
  - $X \sim Gam(n, \lambda) \quad E[X] = \frac{n}{\lambda}$
    - Because $X = \sum_{i=1}^{n} X_i$ for $X_1, \ldots, X_n \sim Exp(\lambda)$

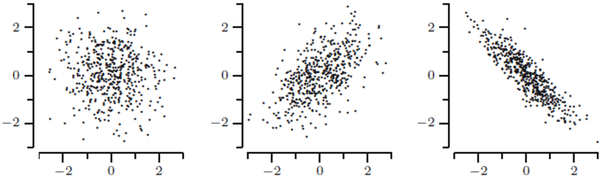
Covariance

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

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

Covariance and correlation coefficient

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.

- Covariance depends on the units of measure!

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

$$-1 \leq \rho(X, Y) \leq 1$$

Sum of independent random variables

- For $X \sim F_X$ and $Y \sim F_Y$, let $Z = X + Y$. We know

$$E[Z] = E[X] + E[Y] \qquad Var(Z) = Var(X) + Var(Y) + 2Cov(X, Y)$$

- What is the distribution function of Z (when X and Y are independent)?
- Examples:
  - For $X \sim Bin(n, p)$ and $Y \sim Bin(m, p)$, $Z \sim Bin(n + m, p)$
  - For $X \sim Geo(p)$ (days radio 1 breaks) and $Y \sim Geo(p)$ (days radio 2 breaks):

$$p_Z(X + Y = k) = \sum_{l=1}^{k-1} p_X(l) \cdot p_Y(k - l) = (k - 1)p^2(1 - p)^{k-2}$$

Adding two independent discrete random variables. Let X and Y be two independent discrete random variables, with probability mass functions p_X and p_Y. Then the probability mass function p_Z of $Z = X + Y$ satisfies

$$p_Z(c) = \sum_j p_X(c - b_j) p_Y(b_j),$$

where the sum runs over all possible values b_j of Y.

Sum of independent random variables

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 $-\infty < z < \infty$.

- The integral is called the **convolution** of $f_X()$ and $f_Y()$
- $X, Y \sim Exp(\lambda)$, $Z = X + Y$, $\quad X, Y, Z \geq 0$ implies $0 \leq Y \leq Z$

$$f_Z(z) = \int_{-\infty}^{\infty} \lambda e^{-\lambda(z-y)} \lambda e^{-\lambda y} dy = \lambda^2 e^{-\lambda z} \int_0^z 1 dy = \lambda^2 e^{-\lambda z} z$$

- $Z = X_1 + \ldots + X_n$ for $X_i \sim Exp(\lambda)$ independent: *[Earlang $Erl(n, \lambda)$ distribution]*

$$f_Z(z) = \frac{\lambda(\lambda z)^{n-1} e^{-\lambda z}}{(n-1)!}$$

$Gam(\alpha, \lambda)$

- Let λ be some average rate of an event, e.g., $\lambda = 1/10$ number of buses in a minute
- The waiting times to see an event is Exponentially distributed. E.g., probability of waiting x minutes to see one bus.
- The waiting times between n occurrences of an event are Erlang distributed. E.g., probability of waiting z minutes to see n buses.

DEFINITION. A continuous random variable X has a *gamma distribution* with parameters $\alpha > 0$ and $\lambda > 0$ if its probability density function f is given by $f(x) = 0$ for $x < 0$ and

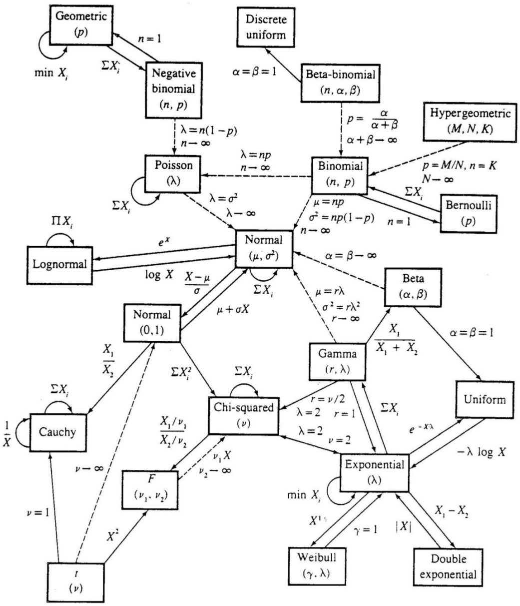
$$f(x) = \frac{\lambda (\lambda x)^{\alpha - 1} e^{-\lambda x}}{\Gamma(\alpha)} \qquad \text{for } x \geq 0,$$

where the quantity $\Gamma(\alpha)$ is a normalizing constant such that f integrates to 1. We denote this distribution by $Gam(\alpha, \lambda)$.

- Extends $Erl(n, \lambda)$ to $\alpha > 0$ by Euler's $\Gamma(\alpha)$

See R script

Common distributions



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).