Beamer presentation using R Markdown

Vincenzo Gioia

08/10/2024

1. Central Limit Theorem (CLT)
2. CLT application to waterpolo goals

Section 1

Central Limit Theorem (CLT)

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots, X_n$, be a sequence of independent and identically distributed (iid) random variables (rv) from a distribution with mean μ and finite variance σ^2.

For large n,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \mathrel{\dot\sim} \mathcal{N}(\mu, \sigma^2/n)$$

where $\dot\sim$ indicates convergence in distribution. Equivalently

$$S_n = \sum_{i=1}^{n} X_i \mathrel{\dot\sim} \mathcal{N}(n\mu, n\sigma^2)$$

The CLT supports the normal approximation to the distribution of a rv that can be viewed as the sum of other rv.

Approximation with CLT: application

The approximation above is useful in statistics for computing some quantities. For instance, let X and Y be two independent Binomial rv, such that $X \sim \mathrm{Bin}(n, p)$ and $Y \sim \mathrm{Bin}(m, q)$.

If we are interested in computing the probability $P(X > Y)$ the Normal approximation is the simplest way to compute it. We can approximate:

$$X \approx \mathcal{N}(np, np(1-p)), \qquad Y \approx \mathcal{N}(mq, mq(1-q)).$$

Then, by using a well known probability result, the difference $W = X - Y$ of two independent normal distributions with means μ_X, μ_Y and variances σ_X^2, σ_Y^2, respectively, is **still a normal distribution** with mean $\mu_W = \mu_X - \mu_Y$ and variance $\sigma_W^2 = \sigma_X^2 + \sigma_Y^2$.

In such a case,

$$\mu_W = \mu_X - \mu_Y = np - mq, \qquad \sigma_W^2 = \sigma_X^2 + \sigma_Y^2 = np(1-p) + mq(1-q).$$

Section 2

CLT application to waterpolo goals

CLT application to waterpolo goals

Tomorrow two professional Italian waterpolo teams, Posillipo and Pro Recco, compete against each other. Let X and Y be the random *goals scored* by Posillipo and Pro Recco, respectively.

We assume that X, Y follow two independent Binomial distributions. Thus, X and Y represent the number of shots converted in goal on the total number of shots n, m made by Posillipo and Pro Recco, with probabilities p and q, respectively.

Before the match, the number of shots is *unknown*. In what follows, we adopt a simplification, and we treat the quantities p, q, m, n as *known*, for instance fixing them upon historical experience:
$p = 0.5, q = 0.7, n = 20, m = 20.$

We want to investigate the Posillipo probability of winning the next match against Pro Recco, that is

$$P(X > Y) = P(X - Y > 0) = ?$$

CLT application to waterpolo goals

So, let W be the r.v., such that $W = X - Y$. We could compute the law of this rv but using the Normal approximation, $W \approx \mathcal{N}(\mu_W = \mu_X - \mu_Y, \sigma^2_W = \sigma^2_X + \sigma^2_Y)$, we can esily compute such a probability of interest.

```
p <- 0.5
q <- 0.7
n <- m <- 20
mW <-  p * n - q * m
sdW <- sqrt( n * p * (1 - p) + m * q * (1 - q) )
# Probability that  W = X-Y > 0 (Posillipo win the match)
PWin_P <- pnorm(0, mean = mW, sd = sdW, lower.tail = FALSE)
PWin_P
```

```
## [1] 0.09362452
```

CLT application to waterpolo goals

Here, we show the probability mass functions of X and Y and the probability density function of $W = X - Y$.

```
# pdf of W
curve(dnorm(x, mW, sdW),
      xlim = c(-12, 20), ylim = c(0, 0.3),
      xlab = "", ylab = "", cex.lab = 1.25)

# pmf of X and Y
points(0 : 20, dbinom(0 : n, n, p), pch = 21, bg = 1, col = "blue")
points(0 : 20, dbinom(0 : m, m, q), pch = 21, bg = 2)

segments(0, -0.01,                 #(x_0, y_0)
         0, dnorm(0, mW, sdW),     #(x_1, y_1)
         lwd = 2)

text(14.25, 0.22, "Y", cex = 2, col = 2)
text(10, 0.2, "X", cex = 2, col = "blue")
text(-4, 0.15, "X - Y", cex = 2, col = 1)
```

CLT application to waterpolo goals