Stochastic Processes

Captain Tsubasa© II: Rise of virtual Champions

José Daniel López-Barrientos, PhD
Facultad de Ciencias Actuariales

Summer 2020

Anáhuac México

Contents

1. I never predict anything, and I never will

Figure 1: Paul Gascoigne


FREUNDSCHAFTSSPIEL
CABJ
BOCA JUNIORS
LA BOMBONERA
ATLÉTICO MADRID

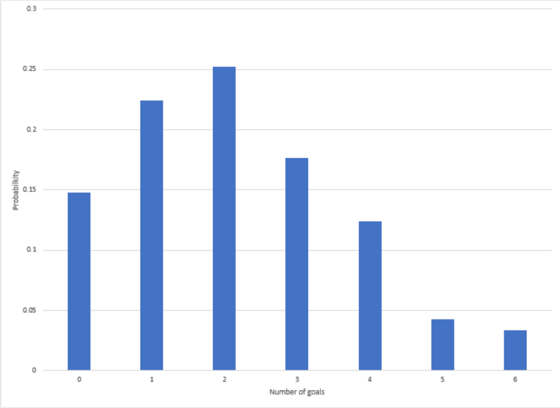

Exhibition
CABJ
CLUB ATLÉTICO DE MADRID
CA BOCA JUNIORS
PES2017

2. A common pattern

La Ligue Sénégalaise de FootBall Professionnel 2017-2018

Statistics	**Numbers**
Minimum	0
Maximum	6
Total goals	455
Total matches	210
Average of goals per match	$2.1\bar{6}$
Probability of goal per minute	2.4%

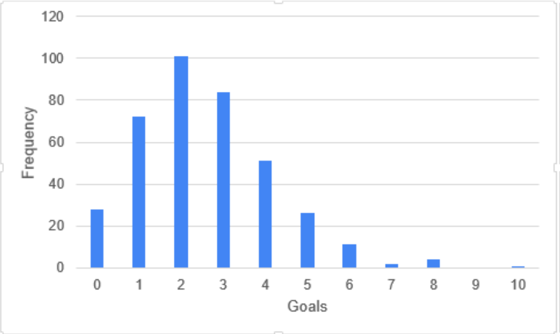
Goal count	Frequency
0	31
1	47
2	53
3	37
4	26
5	9
6	7
Total matches	**455**



LaLiga 2018-2019

Statistics	Numbers
Minimum	0
Maximum	10
Total goals	982
Total matches	380
Average of goals per match	2.58
Probability of goal per minute	2.87%

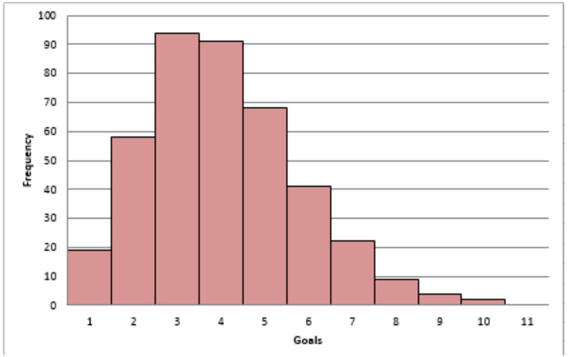
Goal count	Frequency
0	28
1	72
2	101
3	84
4	51
5	26
6	11
7	2
8	4
9	1
10	1
Total matches	**982**



MLS 2019

Statistics	Numbers
Minimum	0
Maximum	9
Total goals	1241
Total matches	408
Average of goals per match	3.04
Probability of goal per minute	3.38%

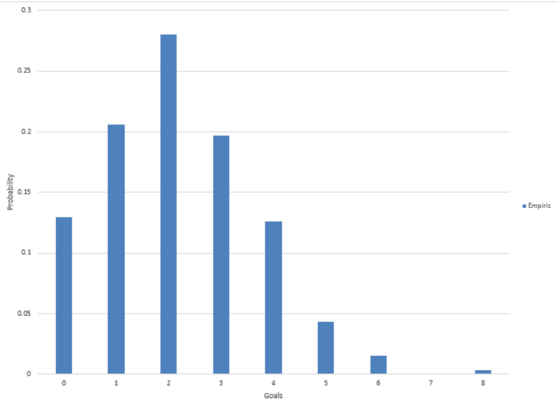
Goal count	Frequency
0	19
1	58
2	94
3	91
4	68
5	41
6	22
7	9
8	4
9	2
Total matches	**408**



Superliga Argentina 2019

Statistics	Numbers
Minimum	0
Maximum	8
Total goals	713
Total matches	325
Average of goals per match	2.19
Probability of goal per minute	2.44%

Goal count	Frequency
0	42
1	67
2	91
3	64
4	41
5	14
6	5
7	0
8	1
Total matches	**325**



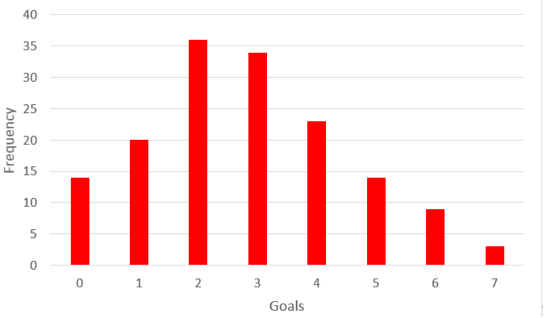
Liga MX Clausura 2016

Statistics	Numbers
Minimum	0
Maximum	7
Total goals	431
Total matches	153
Average of goals per match	2.82
Probability of goal per minute	3.13%

Home Page
Title Page
Contents

Go Back
Full Screen
Close
Quit

Goal count	Frequency
0	14
1	20
2	36
3	34
4	23
5	14
6	9
7	3
Total matches	**153**



3. The Senegalese, Spanish, American, Argentinian, and Mexican connection

Definition 3.1. *We say that a random variable N has a* Binomial distribution *with parameters (n, p) if*

$$\mathbb{P}(N = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{3.1}$$

for $k = 0, 1, ..., n$.

So, we take $n = 90$ "experiments", and p according to:

Country	Senegal	Spain	USA	Argentina	Mexico
Probability	2.4%	2.87%	3.38%	2.44%	3.13%

The tables above should give us the idea of counting the goals in a match. For this purpose, we introduce the following definition.

Definition 3.2. *Let $0 \leq t \leq 1$ be an instant in time, and $N(t)$ be the number of goals scored until moment t of a football match*[1]. *The sequence of random variables $\{N(t) : t \geq 0\}$ is a* ***counting process*** *if it satisfies the following properties:*

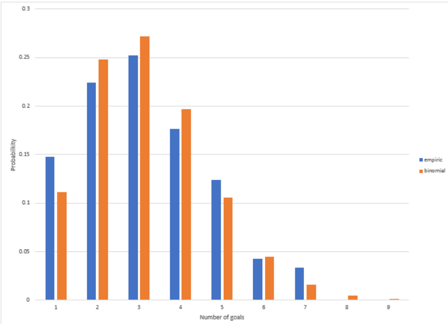
(i) $N(t) \geq 0$,

(ii) *$N(t)$ is valued in the integer numbers,*

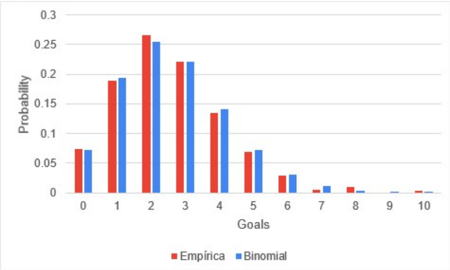
(iii) *If $s < t$, then $N(s) \leq N(t)$,*

(iv) *For $s < t$, $N(t) - N(s)$ is the number of goals scored in the interval $]s;t]$.*

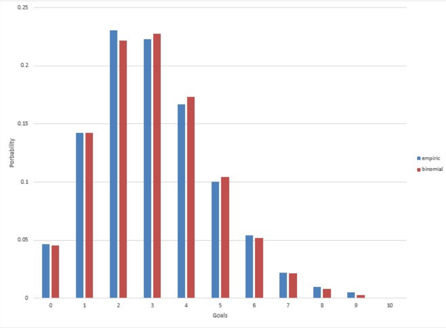
[1] Here, $N(1)$ refers to the total amount of goals scored during the full length of the match, and for example $N\left(\frac{1}{90}\right)$ stands for the total amount of goals scored up to the instant right after the first minuted has passed by.



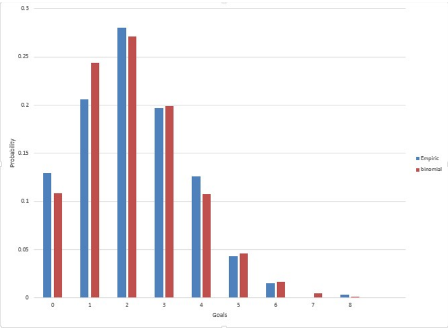
Senegal



Spain



USA



Argentina

4. Only 90 chances to score...

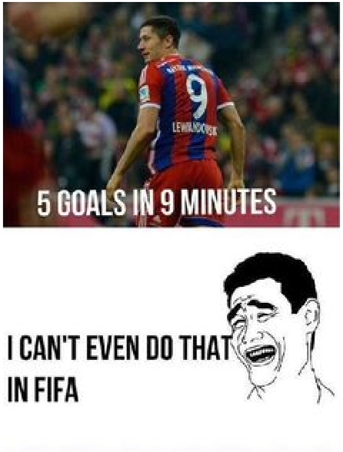


Figure 3: What do we do with him?!

Definition 4.1. *(See [9, Chapter 1].) We say that a counting process* $\{N(t) : t \geq 0\}$ *is a* ***Poisson process with intensity/rate*** $\lambda > 0$ *if:*

(i) $N(0) = 0$,

(ii) *The number of goals scored in different intervals are pairwise independent,*

(iii) *The distributions of* $N(t+s) - N(t)$ *for all* $t \geq 0$ *are equal,*

(iv) *The probability that a single goal is scored in an infinitely small interval of size* h *is*

$$\lim_{h \to 0} \frac{\mathbb{P}[N(h) = 1]}{h} = \lambda, \tag{4.1}$$

(v) *The probability that at least two goals are scored in an infinitely small interval of size* h *is around* 0*, i.e.*

$$\lim_{h \to 0} \frac{\mathbb{P}[N(h) \geq 2]}{h} = 0. \tag{4.2}$$

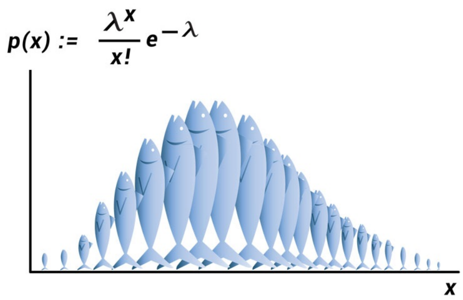


Figure 4: Si vous ne rigolez pas maintenant, alors vous aurez réussi l'examen «Je ne suis pas un nerd».

Theorem 4.2. *Let $N_n(t)$ be a random variable with Binomial distribution and parameters $\left(n, \frac{\lambda t}{n}\right)$, where $\lambda > 0$. Then*

$$\lim_{n\to\infty} \mathbb{P}(N_n(t) = i) = \frac{(\lambda t)^i}{i!} \mathrm{e}^{-\lambda t} \text{ for i=0,1,...} \tag{4.3}$$

The mass function of a Poisson random variable looks like Figure 4. Theorem 4.2 can be found, for instance in [1, p. 204-206], and is often called *Poisson's limit law*. To prove this result, we need the following Lemma.

Lemma 4.3. *If $i \in \mathbb{Z}$, then*

$$\lim_{n\to\infty} \frac{n(n-1)\ldots(n-i+1)}{n^i} = 1.$$

Proof. Note that

$$\begin{aligned}
\lim_{n\to\infty} \frac{n(n-1)\cdots(n-i+1)}{n^i} &= \lim_{n\to\infty} \frac{n}{n}\frac{n-1}{n}\frac{n-2}{n}\cdots\frac{n-i+1}{n} \\
&= \lim_{n\to\infty} 1\left(\frac{n}{n}-\frac{1}{n}\right)\cdots\left(\frac{n}{n}-\frac{i+1}{n}\right) \\
&= 1.
\end{aligned}$$

This proves the result. □

Proof of Theorem 4.2. If $N_n(t)$ follows the Binomial law with parameters $\left(n, \frac{\lambda t}{n}\right)$ then,

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(N_n(t) = i) &= \lim_{n\to\infty} \binom{n}{i} \left(\frac{\lambda t}{n}\right)^i \left(1 - \frac{\lambda t}{n}\right)^{n-i} \\
&= \lim_{n\to\infty} \frac{n!}{(n-i)!i!} \cdot \left(\frac{\lambda t}{n}\right)^i \cdot \left(1 - \frac{\lambda t}{n}\right)^{n-i} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-i+1)}{i!} \frac{(\lambda t)^i}{n^i} \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^i} \\
&= e^{-\lambda t} \frac{(\lambda t)^i}{i!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-i+1)}{n^i}.
\end{aligned}
$$

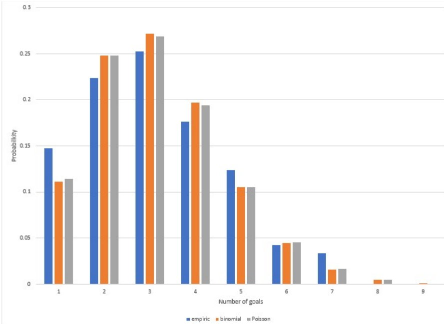
The last equlity follows from the definition of the exponential function, and from the fact that $\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^i = 1$. Finally, Lemma 4.3 yields the result. □

Proof of Theorem 4.2. If $X_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right)$ then,

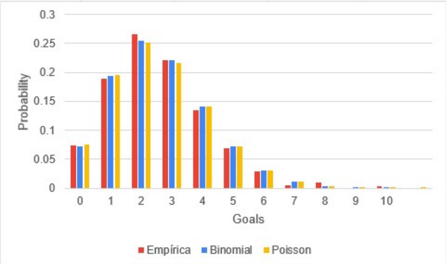
$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X_n = i) &= \lim_{n\to\infty} \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \lim_{n\to\infty} \frac{n!}{(n-i)!i!} \cdot \left(\frac{\lambda}{n}\right)^i \cdot \left(1 - \frac{\lambda}{n}\right)^{n-i} \\
&= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-i+1)}{i!} \frac{\lambda^i}{n^i} \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^i} \\
&= \mathrm{e}^{-\lambda} \frac{\lambda^i}{i!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-i+1)}{n^i}
\end{aligned}$$

The last equality follows from the definition of the exponential function, and from the fact that $\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^i = 1$.. Finally, Lemma 4.3 yields the result. $\square$

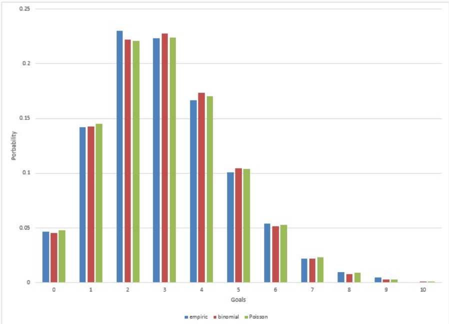
Using $N_\infty(1)$, instead of $N_{90}(1)$ (that is, use formula (4.3) *au lieu de* formula (3.1)), we obtain the following graphs.



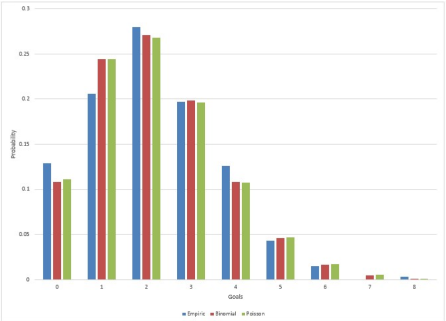
Senegal



Spain



USA



Argentina

Home Page
Title Page
Contents
Page 24 of 40
Go Back
Full Screen
Close
Quit

What have we earned? For starters, we have managed to characterize the whole set of empiric probabilities by a single *Poisson* parameter (instead of the couple of Binomial parameters). Moreover, in view of Theorem 4.2, we can compute the probability that the i-th goal is scored before time t. This is the spirit behind Definition 4.4, and Theorem 4.5.

Definition 4.4. *For a Poisson process, let τ_1 be the time elapsed until the first goal is scored; and τ_i be the time elapsed between the $(i-1)$-th, and the i-th goals, for $i = 2, 3, \dots$ The sequence $\{\tau_i : i = 1, 2, \dots\}$ is called* ***sequence of times of arrival****.*

Theorem 4.5. *(Cf. [3, Theorem 6.8.10].) The random variable τ_i is such that*

$$\mathbb{P}(\tau_i \leq t) = 1 - \mathrm{e}^{-\lambda t} \text{ for } i = 1, 2, \dots, \tag{4.4}$$

and does not depend on τ_j for $i \neq j$ and $j = 1, 2, \dots$

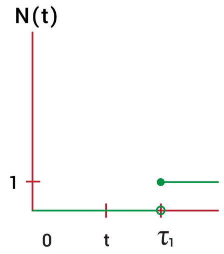


Figure 6: The events $[\tau_1 > t]$ and $[N(t) = 0]$ are equivalent.

Proof. Look at Figure 6 and note that

$$\begin{aligned} \mathbb{P}(\tau_1 > t) &= \mathbb{P}(N(t) = 0) \\ &= e^{-\lambda t}. \end{aligned}$$

Then (4.4) is true for $i = 1$.

Let $[\tau_i|\tau_{i-1} = s]$ be the random variable that represents the time of arrival of the i-th goal given that the $(i-1)$-th goal was scored at time s, for $i = 2, 3, \ldots$. Figure 7 shows that the second goal is scored *only after* time $s + t$. This fact implies that $\mathbb{P}(\tau_2 > t|\tau_1 = s) = \mathbb{P}(N(s+t) - N(s) = 0|N(s) = 1)$.

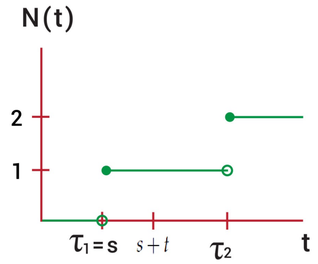


Figure 7: How many goals between s and $s + t$? The answer is zero.

In general, one can argue that

$$
\begin{aligned}
\mathbb{P}(\tau_i > t | \tau_{i-1} = s) &= \mathbb{P}(N(s+t) - N(s) = 0 | N(s) \geq 1) \\
&= \mathbb{P}(N(s+t) - N(s) = 0) \\
&= \mathrm{e}^{-\lambda t}.
\end{aligned}
$$

The second equality follows from Definition 4.1(ii), *i.e.*, the property of independent increments of the Poisson process. It is obvious that τ_i meets (4.4) for $i = 1, 2, \ldots$; moreover, by Definition 4.1(ii), we know that τ_i does not depend on τ_j for $i \neq j$ and $j = 1, 2, \ldots$ □

5. Simulating goals

To generate the first n goals and their times of arrival, we will use Theorem 4.5 and the following result.

Proposition 5.1. *(Cf. [8, Chapter 5.1].) Let U be a uniform random variable on $]0;1[$. For some continuous and invertible distribution function F, the random variable $X := F^{-1}(U)$ has distribution F.*

Proof. Let F_X be the distribution function of $X = F^{-1}(U)$. Then

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \mathbb{P}(F^{-1}(U) \leq x). \end{aligned}$$

The fact that F is a distribution function itself, means that F is monotonic and non-decreasing. So

$$\begin{aligned} F_X(x) &= \mathbb{P}(F^{-1}(U) \leq x) \\ &= \mathbb{P}\left[F(F^{-1}(U)) \leq F(x)\right] \\ &= \mathbb{P}\left[U \leq F(x)\right] \\ &= F(x). \end{aligned}$$

The last equality holds because U is a uniform random variable on $]0;1[$. $\square$

We use Theorem 4.5, and borrow the following algorithm from [8, Chapter 5.4] to simulate the number of goals.

Algorithm 1: Generation of a Poisson realization

Data: Goal average of a team $\lambda > 0$

Result: A simulated number of goals for the team

Generate a random number U;

$N \leftarrow 0$;

$p \leftarrow \exp(-\lambda)$;

$F \leftarrow p$;

while $U \geq F$ **do**

  $p \leftarrow \lambda \frac{p}{N+1}$;

  $F \leftarrow F + p$;

  $N \leftarrow N + 1$;

end

return N;

Now we return to the question from Section 1. To make a proper selection of the goals for and against both teams, we use our information on Atlético de Madrid and Boca Juniors to compile the following table.

Team	Atlético de Madrid	Boca Juniors
Matches at home	19	12
Goals for at home	32	22
Goals against at home	10	7
Matches away	19	13
Goals for away	23	20
Goals against away	19	11

The mean of the goals scored by Atlético de Madrid (A) as host was of $\frac{32}{19}$, whereas Boca Juniors (B) received an average of $\frac{11}{13}$ as visitors. This means that it is reasonable to assume that A will score a mean of

$$\lambda_A = 0.5 \times \frac{32}{19} + 0.5 \times \frac{11}{13} \approx 1.27$$

goals in a direct match in Madrid. Likewise, we assume that

$$\lambda_B = 0.5 \times \frac{20}{13} + 0.5 \times \frac{10}{19} \approx 1.03.$$

We feed Algorithm 1 with these parameters to simulate the match, and obtain...

The mean of the goals scored by B as host was of $\frac{22}{12}$, whereas A received an average of $\frac{19}{19} = 1$ as visitors. This means that it is reasonable to assume that B will score a mean of

$$\lambda_B = 0.5 \times \frac{22}{12} + 0.5 \times 1 \approx 1.42$$

goals in a direct match in Buenos Aires. Likewise, we assume that

$$\lambda_A = 0.5 \times \frac{23}{19} + 0.5 \times \frac{7}{12} \approx 0.90.$$

We feed Algorithm 1 with these parameters to simulate the match, and obtain...

6. A game for the crowds

Theorem 6.1. ***(Weak Law of Large Numbers; Golden Theorem; Bernoulli's Theorem. See [3, Chapters 7.4 and 7.5], [4, Chapter 8], [5, Chapter 8.4], [6, Chapter 8] and [7].)*** *Let $I_1, I_2, \ldots$ be a sequence of independent and identically distributed random variables with common mean μ; and $\bar{I}_n := \frac{I_1+\cdots+I_n}{n}$. Then for any fixed positive number ε,*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{I}_n - \mu\right| > \varepsilon\right) = 0.$$

Proof of Theorem 6.1. Since ε is fixed, then for any $\ell > 0$, we have $\ell\sigma\sqrt{n} < \varepsilon n$ for all sufficiently large values of n. This yields the equivalence of the events

$$\left[\left|\bar{I}_n - \mu\right| > \varepsilon\right] \text{ and } \left[\left|\frac{n\bar{I}_n - n\mu}{\sigma\sqrt{n}}\right| > \ell\right],$$

and so

$$\mathbb{P}\left[\left|\bar{I}_n - \mu\right| < \varepsilon\right] \geq \mathbb{P}\left[\left|\frac{n\bar{I}_n - n\mu}{\sigma\sqrt{n}}\right| < \ell\right] \to \frac{1}{\sqrt{2\pi}}\int_{-\ell}^{\ell} e^{-\frac{x^2}{2}}\,dx \text{ as } n \to \infty. \tag{6.1}$$

The right-most part of (6.1) holds by virtue of the Central Limit Theorem (see [4, Chapter 10.3] and [2, p. 233]). Given any $\delta > 0$ we can first choose ℓ so large that the value of the integral above exceeds $1-\delta$, then choose n so large that (6.1) holds. It then follows that

$$\mathbb{P}\left[\left|\bar{I}_n - \mu\right| < \varepsilon\right] > 1 - \delta$$

for all sufficiently large n. This proves the result. □

Virtual Champions

We used Theorem 6.1, simulated each tournament 10,000 times, and got:

League	3rd place	2nd place	1st place
Superliga Argentina	Boca Juniors	River Plate	**Racing Club**
LaLiga	Real Madrid	Atlético de	**Barcelona**
MLS	Atlanta	New York	**Los Angeles**
LSFP	Teungueth FC	ASC Jaraaf	**Génération Foot**
Liga MX	Tigres	América	**Cruz Azul**

References

[1] Chung, K. L., and AitSahlia, F. *Elementary Probability Theory with Stochastic Processes and an Introduction to Mathematical Finance*. Springer, New York, NY, 2003.

[2] Gnedenko, B. V., and Kolmogorov, A. N. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, 1968.

[3] Grimmett, G., and Stirzaker, D. *Probability and random processes*. Oxford Science Publications, 1994.

[4] Grinstead, C. M., and Snell, J. L. *Introduction to Probability*. American Mathematical Society, 1997.

[5] Hoel, P. G., Port, S. C., and Stone, C. J. *Introduction to Stochastic Processes*. Houghton Miffley Company, Boston, 1972.

[6] Isaac, R. *The Pleasures of Probability*. Springer-Verlag, New York, 1995.

[7] Lanier, D., and Trotoux, D. La loi des grands nombres, le théorème de De Moivre-Laplace. In *Contribution à une approche historique de l'enseignement des mathématiques*, B. Evelyne, M. Claude, and C. FranÃ§ois, Eds. Presses universitaires de Franche-Comté, BesanÃ§on, 1996, pp. 259–294.

[8] Ross, S. *Simulation*. Academic Press, 2012.

[9] Sumpter, D. *Soccermatics: Mathematical Adventures in the Beautiful Game Pro-Edition*. Bloomsbury Sigma, 2017.

Thank you for your attention!

Contact information

Facultad de Ciencias Actuariales, Universidad Anáhuac México.

Av. Universidad Anáhuac 46, Lomas Anáhuac, CP52786

Naucalpan de Juárez, México.

Tel. 01-52-55-56-27-0210 ext. 8506

daniel.lopez@anahuac.mx

https://www.researchgate.net/profile/Jose_Daniel_Lopez-Barrientos

Home Page

Title Page

Contents

Page 40 of 40

Go Back

Full Screen

Close

Quit