Probability

Brief summary

David Manresa

Contents

Chapter 1

The probability model

1.1 Probability space

Definition 1.1. ***Probability space***
A probability space is measure space $(\Omega, \mathscr{A}, P)$*, where*

1. Ω *is the set of all possible outcomes called sample space.*
2. $\mathscr{A}$ *is a* σ*-algebra of* Ω*, called event space.*
3. P *is a probability function* $P : \mathscr{A} \to [0, 1]$*.*

In adition, the probability space must satisfy the three Kolmogorov axioms, that is

1. $P(E) \geq 0 \quad \forall E \in \mathscr{A}$*.*
2. $P(\Omega) = 1$*.*
3. P *must be a* σ*-additivity measure, that is, for any countable family of disjoint sets* $\{E_i\}_{i\in I}$

$$P\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} P(E_i).$$

Properties

1. Let $A, B \in \mathscr{A}$, $A \subset B$, then $P(A) \leq P(B)$ and $P(B \backslash A) = P(B) - P(A)$.
2. $P(A^c) = 1 - P(A) \quad \forall A \in \mathscr{A}$.
3. $P(\emptyset) = 0$.
4. Let $A_1, \ldots, A_n \in \mathscr{A}$ such that $A_i \cap A_j = \emptyset \quad \forall i \neq j$, then $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \ldots P(A_n)$.

1.2 Inclusion-exclusion principle

Theorem 1.1. ***Inclusion-exclusion principle***
Let $A_1, A_2, \dots, A_n$ be events of a probability space $(\Omega, \mathscr{A}, P)$, then

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} \sum_{1\leq \dots \leq n} P\Big(\bigcap_{i=1}^{n} A_i\Big),$$

or equivalently

$$P\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} \Big(\sum_{1\leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k})\Big).$$

1.3 Computing discrete probabilities

We compute discrete probabilities using the ratio of favorable outcomes to the total number of outcomes. To count these outcomes, we can use various combinatorial formulas.

Combinations
The number of ways to chose k elements from a set of n elements (ignoring order) is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Combinations with repetition
The number of ways to chose k elements with repetition from a set of n elements (ignoring order) is

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}.$$

Permutations
The number of ways to arrange n elements (not ignoring order) is

$$n! = n(n-1)(n-2)\dots 1.$$

Permutations with repetition
The number of ways to arrange r elements from a set of n elements with repetition (not ignoring order) is

$$n^r.$$

Permutations with identical objects
The number of ways to arrange n elements where there are $n_1, \dots, n_k$ that are identical is

$$\frac{n!}{n_1! \dots n_k!}.$$

1.4 Computing continuous probabilities

Plenty of problems can be modeled with the probability space $(\Omega, \mathbb{B}_\Omega, P)$, where $\Omega \subset \mathbb{R}^k$, $\mathbb{B}_\Omega$ is the minimum σ-algebra on Ω, and $P(A) = \dfrac{\lambda_k(A \cap \Omega)}{\lambda_k(\Omega)}$, where $\lambda_k : \mathbb{B}^k \to \mathbb{R}$ is the Lebesegue measure on $\mathbb{B}_{\mathbb{R}^k} = \mathbb{B}^k$.

Therefore, the problem of finding the probability of some event $A \in \mathbb{B}_\Omega$ lies on computing the "length" (the respective hypervolume depending the dimension of the sample space) of A and Ω.
For example, if our sample space is the square $[-1,1] \times [-1,1]$ and we want to compute the probability of choosing a random point that lies on the unit circle, the probability is just $\dfrac{\pi}{4}$.

This theoretical approach to compute continuous probabilities will be simplified with random variables and their density distributions. For example, the probability that we just computed, is just calculating the probability of a uniform random variable on $[-1,1] \times [-1,1]$ landing on the unit circle.

1.5 Exercises

1. Model the probability space of extracting a ball from an urn that containts 3 balls.

Numbering each ball, the posible events are then $\Omega = \{1, 2, 3\}$.

The σ-algebra in this case is just the power set $\mathscr{P}(\Omega)$

$$\mathscr{P}(\Omega) = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}, \emptyset\}.$$

We can construct the probability function $P : \mathscr{P}(\Omega) \to [0,1]$ assuming each ball has the same probability to get extracted, then

$P(\{1\}) = P(\{2\}) = P(\{3\}) = 1/3,$
$P(\{1,2\}) = P(\{1,3\}) = P(\{2,3\}) = 2/3,$
$P(\{1,2,3\}) = 1,$
$P(\emptyset) = 0.$

Then $(\Omega, \mathscr{P}(\Omega), P)$ is a probability space. (It satisfies the first and second Kolmogorov axioms trivially and it's easy to check the third).

2. n cards numbered from 1 to n are shuffled and placed in a line on the table. A match occurs at position i if the card in that position has the number i. What is the probability that at least one match occurs?

Let A_i be the event that there is a coincidence in the i-th position. The total number of posible permutations of the cards are $n!$, while if there is a coincidence in the i-th position, the total number of favorable permutations are $(n-1)!$ ($i - th$ fixed and we permute the rest of the cards). Then

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Anagously, fixing the $i_1, \ldots, i_k$ positions

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{(n-k)!}{n!}.$$

Then, by the inclusion-exclusion principle

$$
\begin{aligned}
P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} \sum_{1<\cdots<n} P\Big(\bigcap_{i=1}^{n} A_i\Big) \\
&= n\frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \binom{n}{3}\frac{1}{n(n-1)(n-2)} \cdots + (-1)^{n+1}\frac{1}{n!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + (-1)^{n+1}\frac{1}{n!}.
\end{aligned}
$$

Chapter 2

Conditional probability

Definition 2.1. ***Conditional probability***
Let A, B sucesses of a probability space $(\Omega, \mathscr{A}, P)$, such that $P(B) > 0$, then the probability of A conditioned by B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

It is useful to compute some probabilities the identity

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}),$$

which can be proven easily by substituing each $P(A_j|A_1 \cap \cdots \cap A_{j-1})$ by $\dfrac{P(A_1 \cap \cdots \cap A_j)}{P(A_1 \cap \cdots \cap A_{j-1})}$, $\forall j = 1, \ldots, n$.

2.1 Law of total probability

Theorem 2.1. ***Law of total probability***
Let $\{B_n\}_{n \in I}$ be a finite or infinitely numerable set of events of a probability space $(\Omega, \mathscr{A}, P)$ such that $B_i \cap B_j = \emptyset$ for all $i \neq j$ and $\bigcup_n B_n = \Omega$, then

$$P(A) = \sum_n P(B_n)P(A|B_n).$$

2.2 Bayes theorem

Theorem 2.2. ***Bayes theorem***
Let A, B events of a probability space $(\Omega, \mathscr{A}, P)$, such that, $P(B) \neq 0$, then

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

Combining with the law of total probability, then

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_n P(A_n)P(B|A_n)},$$

where $\{A_n\}_{n\in I}$ *is a set of events that has the properties to hold the law of total probability.*

2.3 Independence of events

Definition 2.2. ***Pairwise independence***
Let $\{A_i\}_{i=1}^n$ *a set of events of* $\mathscr{A}$*, then the events are pairwise independent if and only if*

$$P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i \neq j.$$

Definition 2.3. ***Mutually independence***
Let $\{A_i\}_{i=1}^n$ *be a set of events of* $\mathscr{A}$*, then the events are mutually independent if and only if*

$$P\Big(\bigcap_{j=1}^{k} A_{i_j}\Big) = \prod_{j=1}^{k} P(A_{i_j}),$$

$\forall k \leq n$ *and for every* $i_1, \ldots, i_k$ *such that* $1 \leq i_1 < \cdots < i_k \leq n$.

2.4 Exercises

1. Compute the probability that n people have distinct birthdays in a N day year.

Applying the identity

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

The first person A_1 can have his birthday in any of the N days. Given the birthday of the A_1, the second person A_2 can only have his birthday in the rest $N-1$ days, then $P(A_1) = \frac{N}{N} = 1$, $P(A_2|A_1) = \frac{N-1}{N}$, and in general, $P(A_j|A_1 \cap \cdots \cap A_{j-1}) = \frac{N-j+1}{N}$, $j = 1, \ldots, n$.

Therefore, the probability is

$$\frac{(N-1)(N-2)\cdots(N-n+1)}{N^{n-1}}.$$

2. If we choose a random point inside the circle of radius 3, compute the probability that the $-1 < y < 1$, given that the point satisfies $x^2 + y^2 = 4$.

Let $\Omega = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 9\}$, $A = \{(x,y) \in \mathbb{R}^2 : -1 < y < 1\}$ and $B = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 4\}$.

Then $P(B) = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{4\pi}{9\pi} = \frac{4}{9}$.

To compute $P(A \cap B)$, notice that if we divide their intersection with the lines from $(-\sqrt{3}, -1)$ to $(\sqrt{3}, 1)$ and from $(-\sqrt{3}, 1)$ to $(\sqrt{3}, -1)$ we obtain two isosceles triangles and two circular sectors. Therefore

$$P(A \cap B) = \frac{\lambda(A \cap B)}{\lambda(\Omega)} = \frac{2\sqrt{3} + \frac{4\pi}{3}}{9\pi}$$

Concluding

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\sqrt{3}}{2\pi} + \frac{1}{3}.$$

3. Some clinical analysis is employed in the diagnosis of three sicknesses, B_1, B_2, and B_3. The proportion of sick people with each of these are 3%, 2%, and 1% respectively. On the other hand, the analysis gives a positive result with probabilities $0.85, 0.92$, and 0.78 for sick people with these three sicknesses, and 0.005 for people who are not sick.

1. What is the probability of obtaining a positive result on a random person?
2. What is the probability that the person has the sickness B_i given that the result is positive?

1. Let A be the event of obtaining a positive result. Notice that the problem gives us $P(A|B_i)$ for $i = 0, \ldots, 3$, where B_0 is the event that the person is not sick. So we just need to compute $P(B_0)$, which is $P(B_0) = 1 - (P(B_1) + P(B_2) + P(B_3)) = 0.94$.

We have that $B_i \cap B_j = \emptyset \ \forall i \neq j$ and $\sum_{i=0}^{3} P(B_i) = 1$,then, applying the law of total probability

$$P(A) = \sum_{i=0}^{3} P(B_i)P(A|B_i),$$

$$P(A) = 0.03 \cdot 0.85 + 0.02 \cdot 0.92 + 0.01 \cdot 0.78 + 0.94 \cdot 0.005 = 0.0564.$$

2. Applying the Bayes theorem

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{0.03 \cdot 0.85}{0.0564} \simeq 0.452,$$

$$P(B_2|A) = \frac{P(A|B_2)P(B_2)}{P(A)} = \frac{0.02 \cdot 0.92}{0.0564} \simeq 0.326,$$

$$P(B_3|A) = \frac{P(A|B_3)P(B_3)}{P(A)} = \frac{0.01 \cdot 0.78}{0.0564} \simeq 0.138.$$

Chapter 3

Random variables

3.1 Random variables and their distributions

Definition 3.1. ***Random variable***
Let $(\Omega, \mathscr{A}, P)$ *be a probability space, a random variable* X *is a measurable function* $X : \Omega \to \mathbb{R}$.

Definition 3.2. ***Distribution function***
Let X *be a random variable on the probability space* $(\Omega, \mathscr{A}, P)$*, the distribution function of the variable* X *is the function* $F_X : \mathbb{R} \to [0, 1]$ *defined as*

$$F_X(x) = P(X \leq x).$$

Definition 3.3. ***Density function***
Let X *be a continuous random variable on the probability space* $(\Omega, \mathscr{A}, P)$*, the density function of the variable* X *is the function* $f_X : \mathbb{R} \to \mathbb{R}$ *defined as*

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} f(x)dx.$$

Notice that if $f_X(x)$ *is continuous, then* $\dfrac{dF_X}{dx}(x) = f_X(x)$

Definition 3.4. ***Expected value***
Let X *be a random variable on a probability space* $(\Omega, \mathscr{A}, P)$*. The expected value of the random variable* X *is defined as*

$$E[X] = \sum_{x_k} x_k P(X = x_k),$$

if it is discrete, and

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

if it is continuous.

Properties

1. $E[c] = c$.

2. $E[c_1X_1 + c_2X_2] = c_1E[X_1] + c_2E[X_2]$.
3. If X es a positive random variable, then $E[X] \geq 0$.
4. $E[I_A] = P(A)$, where I_A is the indicator function of the event A.
5. Let X be a positive integer random variable, then $E[X] = \sum_{m=1}^{\infty} P\{X \geq m\}$.

3.2 Raw and central moments

Definition 3.5. ***r-th raw moment (α_r)***
The $r-th$ moment of a random variable X (or its distribution), is the expected value of X^r, that is

$$E[X^r] = \sum_{x_k} x_k^r P(X = x_k),$$

if it is discrete, and

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

if it is continuous.

Definition 3.6. ***r-th central moment (μ_r)***
The $r-th$ moment of a random variable X (or its distribution) is de expected value of $(X - E([X])^r$, that is

$$E[(X - E[X])^r] = \sum_{x_k} (x_k - E[X])^r P(X = x_k),$$

if it is discrete, and

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

if it is continuous.

It is possible to establish a relation between raw moments and central moments as

$$\mu_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \alpha_1^k \alpha_{r-k}.$$

The expected value (mean) and the second central moment (variance) are the most common and have special notations, μ and σ^2 respectively.

3.3 Joint probability space

Definition 3.7. ***Joint probability space***
Let $(\Omega_i, \mathscr{A}_i, P_i)$ $i = 1, \ldots, n$ be probability spaces, then the joint space

$$(\Omega_1 \times \cdots \times \Omega_n, \mathscr{A}_i \otimes \cdots \otimes \mathscr{A}_n, P)$$

is a probability space, where $P(A_1 \times \cdots \times A_n) = P_1(A_1) \cdots P_n(A_n)$ $\forall A_i \in \mathscr{A}_i$.

In practice, we can compute the probability of a repeated experiment of a discrete random variable constructing the required joint probability space with itself. Since each experiment is independent from the previous one, then trivially

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

Definition 3.8. ***Marginal distribution***
Let $(\Omega_1 \times \Omega_2, \mathscr{A} \otimes \mathscr{A}, P)$ be a joint probability space and let $X = (X_1, X_2)$ be a random variable on it, then the marginal distributions of X_1 and X_2 are

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2).$$

$$P(X_2 = x_2) = \sum_{x_1} P(X_1 = x_1, X_2 = x_2),$$

if they are discrete, and

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2,$$

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1.$$

if they are continuous.

Definition 3.9. ***Conditional distribution***
Let X_1 and X_2 be random variables on a probability space $(\Omega, \mathscr{A}, P)$. Then the conditional probability distribution is

$$P(X_1 \leq x_1 | X_2 = x_2) = \frac{P(X_1 \leq x_1, X_2 = x_2)}{P(X_2 = x_2)}$$

if they are discrete, or

$$f(x_1 | x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}$$

if they are continuous.

Definition 3.10. ***Convolution***
Let X_1 and X_2 be random variables in a probability space $(\Omega, \mathscr{A}, P)$, the random variable $X_1 + X_2$ has distribution

$$P(X_1 + X_2 = s) = \sum_{x_1} P(X_1 = x_1, X_2 = s - x_1),$$

if they are discrete, and

$$(f_{X_1} * f_{X_2})(s) = f_{X_1 + X_2}(s) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(s - x) dx_1.$$

if they are continuous.

3.4 Independence of random variables

Definition 3.11. ***Independence***
Let $X_1, \ldots, X_n$ be random variables on the same probability space $(\Omega, \mathscr{A}, P)$. Then, these random vari-

ables are independent if and only if

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = F_1(x_1)\cdots F_n(x_n) \quad \forall x_1, \cdots x_n \in \mathbb{R},$$

where F_i is the distribution function of X_i.

3.5 Exercises

1. A coin has a probability p of landing heads. Determine the distribution of the number of flips until the $i-th$ head appears.

Let X be the random variable of the number of flips until the $i-th$ head appears. We have to compute $P(X=k)$. Once we know that the $k-th$ flip is the last head, we need to distribute $i-1$ heads among the first $k-1$ flips, there are $\binom{k-1}{i-1}$ ways to do this, where we have $i-1$ heads and $(k-1)-(i-1)=k-i$ tails. The probability of this arrangement is then $\binom{k-1}{i-1}(1-p)^{k-i}p^{i-1}$, so multiplying by p (the last head) we obtain

$$P(X=k) = \binom{k-1}{i-1}(1-p)^{k-i}p^i, \quad k=i,i+1,\dots.$$

2. An urn contains balls numbered from 1 to n, r draws are made. Find the probability that the maximum result is k if the draws are made with replacement and without replacement.

Let M be the maximum number of the r extractions. We will compute $P(M=k)=P(M\leq k)-P(M\leq k-1)$. If the draws are made with replacement, the total amount of cases are n^r, because in each extraction there are n possible balls, and we repeat that r times. The favorable cases are those when we extract a ball with a number less or equal than k, again, we repeat that r times, so we have k^r favorable cases. Then

$$P(M\leq k) = \left(\frac{k}{n}\right)^r,$$

therefore

$$P(M=k) = P(M\leq k) - P(M\leq k-1) = \left(\frac{k}{n}\right)^r - \left(\frac{k-1}{n}\right)^r.$$

If the draws are made without replacement, the total amount of cases are $\binom{n}{r}$ while the favorable ones are $\binom{k}{r}$, then

$$P(M=k) = \frac{\binom{k}{r}-\binom{k-1}{r}}{\binom{n}{r}} = \frac{\binom{k-1}{r-1}}{\binom{n}{r}}$$

3. Let (X,Y) be a random variable with density distribution

$$f(x,y) = kx^2y \quad 0<y<x<1.$$

Compute k and the marginal and conditional distributions.

In order to be a density distribution it must satisfy

$$1 = \int_0^1 \int_0^x kx^2 y dy dx = \frac{k}{10}$$

Then $k = 10$.

The marginal density functions are

$$f_X(x) = \int_0^x 10x^2 y dy = 5x^4,$$

$$f_Y(y) = \int_y^1 10x^2 y dx = \frac{10}{3}(y(1-y^3).$$

While the conditional density functions are

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{10x^2 y}{10y(1-y^3)/3} = \frac{3x^2}{1-y^3}, \quad y < x < 1.$$

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{10x^2 y}{5x^4} = \frac{2y}{x^2}, \quad 0 < y < x.$$

Chapter 4

Discrete probability distributions

4.1 Bernoulli distribution

The proability distribution that takes the value 1 with probability p and the value 0 with probability $1-p$.

$$P(X=k) = \begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0. \end{cases}$$

$$E[X] = p.$$
$$\sigma^2(X) = p(1-p).$$

4.2 Binomial distribution

The probability distribution of the number X of successes in n Bernoulli trials.

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E[X] = np.$$
$$\sigma^2(X) = np(1-p).$$

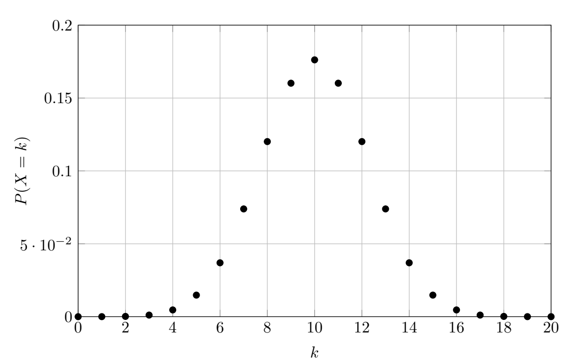

Figure 4.1: Binomial distribution, $n = 20$, $p = 0.5$.

4.3 Negative binomial distribution

The probability distribution of the number of failures X before r successes.

$$P(X = k) = \binom{k + r - 1}{k}(1 - p)^k p^r$$

$$E[X] = \frac{r(1 - p)}{p}$$

$$\sigma^2(X) = \frac{r(1 - p)}{p^2}$$

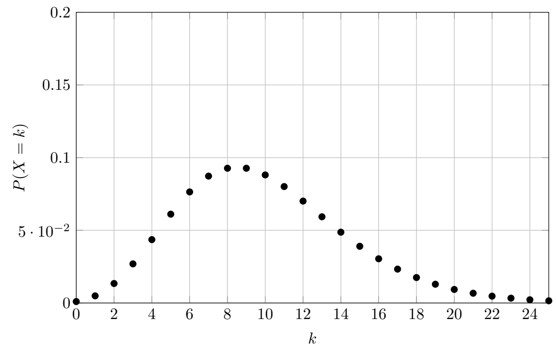

Figure 4.2: Negative Binomial distribution, $r = 10$, $p = 0.5$.

4.4 Poisson distribution

The probability distribution of the number of events X occurring in a fixed interval of time if these events occur with a known constant mean rate λ and independently of the time since the last event.

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$E[X] = \lambda$$

$$\sigma^2(A) = \lambda$$

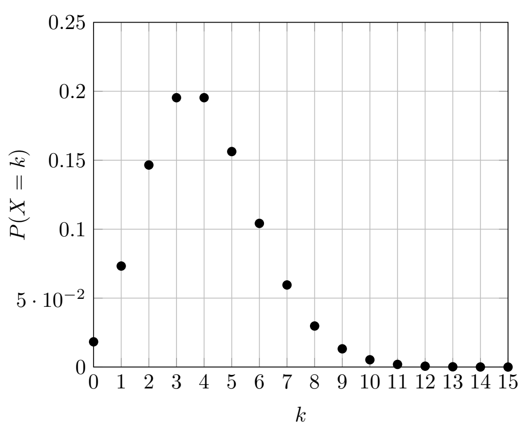

Figure 4.3: Poisson distribution, $\lambda = 4$

4.5 Geometric distribution

The probability distribution of the number X of Bernoulli trials needed to get one success

$$P(X = k) = (1 - p)^{k-1} p$$

$$E[X] = \frac{1}{p},$$
$$\sigma^2(X) = \frac{1-p}{p^2}.$$

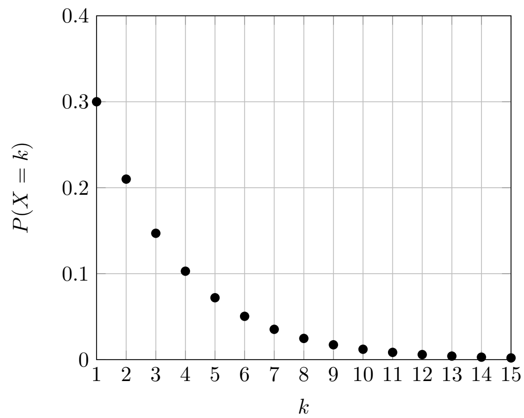

Figure 4.4: Geometric distribution $p = 0.3$

4.6 Hipergeometric distribution

The distribution of k successes in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature.

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$E[X] = n\frac{K}{N}$$

$$\sigma^2(X) = n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}$$

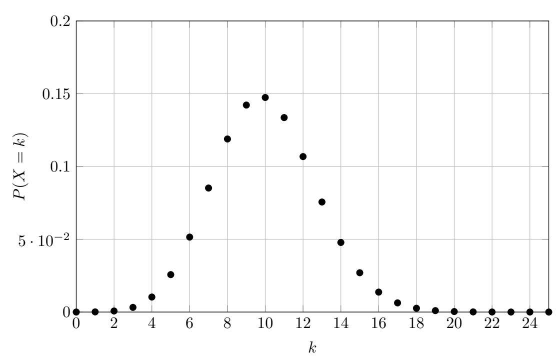

Figure 4.5: Hypergeometric distribution, $N = 500$, $K = 50$, $n = 100$.

Chapter 5

Continuous probability distributions

5.1 Normal distribution $N(\mu, \sigma)$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$E[X] = \mu,$$
$$Var[X] = \sigma^2.$$

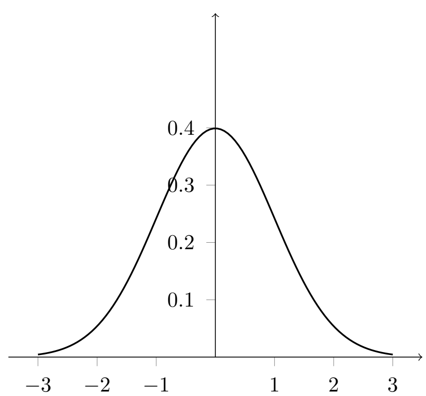

Figure 5.1: Standard Normal Distribution, $N(0, 1)$.

5.2 Uniform distribution $U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise.} \end{cases}$$

$$E[X] = \frac{1}{2}(a+b),$$
$$Var[X] = \frac{1}{12}(b-a)^2.$$

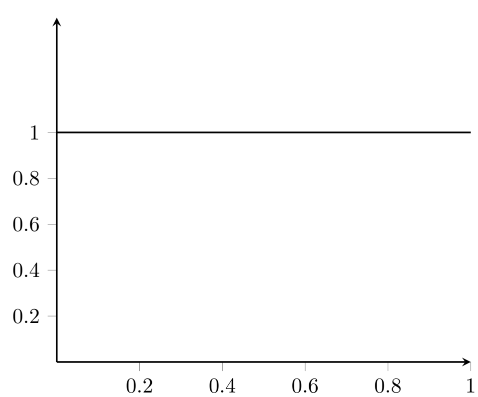

Figure 5.2: Uniform Distribution $U(0, 1)$

5.3 Exponential distribution $exp(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & if\, x < 0. \end{cases}$$

$$E[X] = \frac{1}{\lambda},$$
$$Var[X] = \frac{1}{\lambda^2}.$$

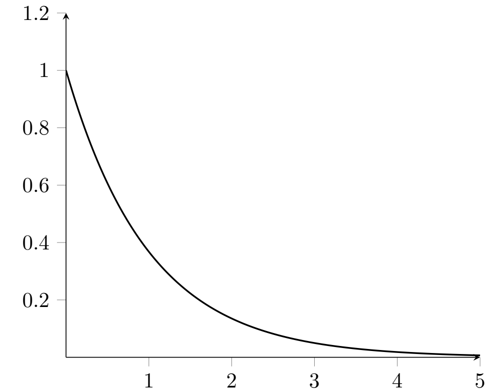

Figure 5.3: Exponential Distribution, $\lambda = 1$

5.4 Gamma distribution $\gamma(k, \lambda)$

$$f(x) = \frac{(\lambda x)^{k-1} \lambda e^{-\lambda x}}{\Gamma(k)} \quad x > 0.$$

$$E[X] = \frac{k}{\lambda},$$
$$Var[X] = \frac{k}{\lambda^2}.$$

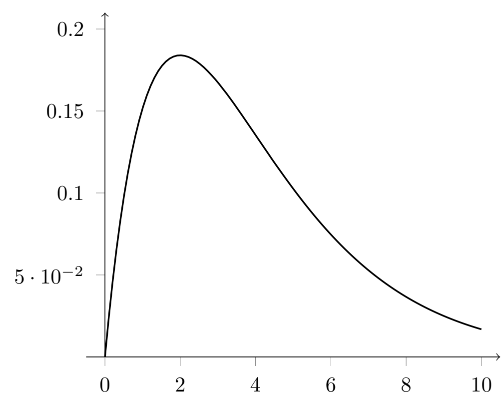

Figure 5.4: Gamma Distribution, $k = 2, \lambda = 0.5$.

5.5 Beta distribution Beta(α, β)

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

$$E[X] = \frac{\alpha}{\alpha+\beta},$$

$$Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

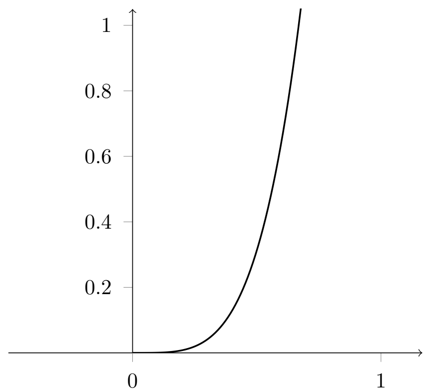

Figure 5.5: Beta Distribution $\alpha = 5, \beta = 1$

Be aware that the shape of the density function of a Beta random variable changes drastically with the parametres α and β.

Chapter 6

Important results

6.1 Random variable convergences

Definition 6.1. ***Convergence in distribution***
Let $\{X_n\}_{n\in\mathbb{N}}$ *be a secuence of random variables defined on a probability space* $(\Omega, \mathscr{A}, P)$ *with distributions* $\{F_n\}_{n\in\mathbb{N}}$*, then the secuence converges in distribution towards a random variable* X *with distribution* F *and it is denoted as* $X_n \xrightarrow{d} X$*, if*

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \forall x \in \mathbb{R}.$$

Definition 6.2. ***Convergence in probability***
Let $\{X_n\}_{n\in\mathbb{N}}$ *be a secuence of random variables defined on a probability space* $(\Omega, \mathscr{A}, P)$*, then the secuence converges in probability towards a random variable* X *and it is denoted as* $X_n \xrightarrow{p} X$*, if for all* $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

Definition 6.3. ***Convergence almost surely***
Let $\{X_n\}_{n\in\mathbb{N}}$ *be a secuence of random variables defined on a probability space* $(\Omega, \mathscr{A}, P)$*, then the secuence converges almost surely towards a random variable* X *and it is denoted as* $X_n \xrightarrow{a.s} X$*, if for all* $\varepsilon > 0$

$$P(\lim_{n\to\infty} X_n = X|) = 1.$$

6.2 Central limit theorem

Theorem 6.1.
Let $X_1, \ldots, X_n$ *independent random variable identically distributed, with finite variance* σ^2 *and mean* μ*. Then* $S_n = X_1 + \cdots + X_n$ *has mean* $n\mu$ *and variance* $\sqrt{n}\sigma^2$*.*

Theorem 6.2. ***Central limit theorem***
Let $X_1, X_2, \ldots$ *be a sequence of independent and identically distributed random variables with finite variance* σ^2 *and mean* μ*, then*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2), \quad as \quad n \to \infty,$$

that is the random variable $\sqrt{n}(\hat{X}_n - \mu)$ *converges in distribution to a normal* $N(0, \sigma^2)$*.*

This theorem can be applied also to the sum S_n of the variables in the sequence, that is

$$S_n = X_1 + \dots X_n \xrightarrow{d} N(n\mu, \sqrt{n}\sigma^2) \ \ as \ \ n \to \infty.$$

Which we obtain the usual central limit theorem for the mean after standardize this expression and dividing by n on the numerator and denominator.

6.3 Weak law of large numbers

Theorem 6.3. ***Weak law of large numbers***
Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables, then the mean $\bar{X}_n = \frac{1}{n}(X_1 + \dots X_n)$ converges in probability to $E[X]$ as $n \to \infty$, that is

$$\bar{X}_n \xrightarrow{p} E[X].$$

6.4 Strong law of large numbers

Theorem 6.4. ***Strong law of large numbers***
Let $X_1, X_2, ...$ be a sequence of independent and identically distributed random variables, then the mean $\bar{X}_n = \frac{1}{n}(X_1 + \dots X_n)$ converges almost surely to $E[X]$ as $n \to \infty$, that is

$$\bar{X}_n \xrightarrow{a.s} E[X] \ \ as \ n \to \infty,$$

References

[1] Ricardo Velez Ibarrola. *Cálculo de Probabilidades II*. UNED, 2019.

[2] Sheldon M.Ross. *Introduction to probability models*. Academic Press, 2019.

[3] Ricardo Velez Ibarrola y Víctor Hernández Morales. *Cálculo de Probabilidades I*. UNED, 2011.