Probability Proofs (ACSAI)

Gianmaria Romano

A.Y. 2024/2025 (First Semester)

Disclaimer

This document contains the proofs given by Professor Bertini for the "Probability" course for the "Applied Computer Science and Artificial Intelligence" course at Sapienza Università di Roma.
Keep in mind, however, that these notes do **not** replace the course material as they are meant to be used for preparation to the oral exam, meaning that it is still suggested to check the Professor's resources as well.
These notes are free to use/share but please remember to credit me as the author and therefore do not hide/remove this page.

Contents

Chapter 1

Conditional probability

1.1 The law of total probability

Given a partitioned sample space $\Omega = \bigcup_{i=1}^{n} D_1$ *and an event* $A \subset \Omega$*, it is possible to state that* $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(D_i)\mathbb{P}(A|D_i)$*.*
For proof, if $A \subset \Omega$, then $A \cap \Omega = A$, but, by definition of Ω, it is possible to apply distributivity and rewrite the statement as:

$$A = A \cap (\bigcup_{i=1}^{n} D_i) = \bigcup_{i=1}^{n} (A \cap D_i)$$

Most particularly, all the sets $A \cap D_i$ are disjoint due to the partition, allowing to apply additivity and state that:

$$\mathbb{P}(A) = \mathbb{P}(\bigcup_{D_i}^{n} (A \cap D_i)) = \sum_{i=1}^{n} \mathbb{P}(A \cap D_i)$$

By applying conditional probability, it is possible to conclude that:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(D_i)\mathbb{P}(A|D_i) \text{ because } \mathbb{P}(A|D_i) = \frac{\mathbb{P}(A \cap D_i)}{\mathbb{P}(D_i)}$$

1.2 Bayes' theorem

Given two events $A, B \subset \Omega$*, it is possible to write* $\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$*.*
For proof, apply the definition of conditional probability:

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

At the same time, however, it is possible to do the same on the "reverse condition":

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

Therefore, it is possible to rewrite the first statement as:

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

Chapter 2

Natural probability distributions

2.1 Sample spaces of probability distributions

2.1.1 Bernoulli distribution

Given $\Omega = \{0, 1\}$, the Bernoulli distribution of the weights $p_1 = p$ (success) and $p_0 = 1 - p$ (failure) is a legit probability distribution.
For proof, apply additivity on the two possible elementary events:

$$\mathbb{P}(\Omega) = p_0 + p_1 = 1 - p + p = 1$$

2.1.2 Binomial distribution

Given $\Omega = \{0, \ldots, n\}$, the binomial distribution of parameters n and p of the weights $p_k = \binom{n}{k} p^k (1-p)^{n-k}$ modelling the probability of seeing k heads in n (biased) coin tosses (with $0 \leq k \leq n$) is a legit probability distribution.
For proof, apply additivity and exploit the Newton binomial:

$$\mathbb{P}(\Omega) = \sum_{k=0}^{n} p_k = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1$$

2.1.3 Geometric distribution

Given $\Omega = \mathbb{N}$, the geometric distribution of parameter p collecting the weights $p_k = (1-p)^{k-1} p$ modelling the probability of seeing head for the first time at the k^{th} trial of a repeated (biased) coin toss is a legit probability distribution.
For proof, apply additivity to exploit the properties of geometric series:

$$\mathbb{P}(\Omega) = \sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=1}^{\infty}(1-p)^{k-1}$$

Notice that the summation is basically a geometric series with common ratio $1-p \in (0,1)$, meaning that the series converges and:

$$\mathbb{P}(\Omega) = p\frac{1}{1-(1-p)} = \frac{p}{p} = 1$$

2.1.4 Poisson distribution

Given $\Omega = \mathbb{Z}^+$, the Poisson distribution of parameter λ with weights $p_k = \frac{e^{-\lambda}\lambda^k}{k!}$ modelling the probability of k successes given an average rate λ of successes is a legit probability distribution.

For proof, apply additivity in order to recover a Taylor expansion:

$$\mathbb{P}(\Omega) = \sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}$$

However, notice that the summation is actually a McLaurin expansion for $f(x) = e^x$, with $x = \lambda$, allowing to state that:

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda} \Rightarrow \mathbb{P}(\Omega) = e^{-\lambda}e^{\lambda} = 1$$

Chapter 3

Random variables

3.1 Variance of a random variable

Given a random variable X with finite expectation $\mathbb{E}(X) = \mu$, its variance is given by $\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}^2(X)$.
For proof, let $\mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}^2(X))$ and, by exploiting the linearity of expectation, rewrite this value as:

$$\mathbb{V}(X) = \mathbb{E}(X^2) + \mathbb{E}(-2X\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}^2(X))$$

However, if $\mathbb{E}(X) = \mu$ is a constant, let $\mathbb{E}(\mu) = \mu$, allowing to simplify:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}^2(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

3.2 Covariance between two random variables

The covariance between two random variables X and Y can be expressed as $Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
For proof, rewrite the formula as:

$$Cov(X,Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY - X\mathbb{E}(Y) - Y\mathbb{E}(X) + \mathbb{E}(X)\mathbb{E}(Y))$$

Most particularly, assuming $\mathbb{E}(X) = \mu_X$ and $\mathbb{E}(Y) = \mu_Y$, it is possible to exploit linearity of the expectation and conclude that:

$$Cov(X,Y) = \mathbb{E}(XY) - \mu_Y\mathbb{E}(X) - \mu_X\mathbb{E}(Y) + \mu_X\mu_Y = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

3.2.1 Independence implies uncorrelation

If X and Y are independent random variables, then $Cov(X,Y) = 0$ and the two variables are said to be uncorrelated.
For proof, remember that, if X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$,

meaning that:

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0$$

Keep in mind however, that, while independence guarantees uncorrelation, uncorrelation does not necessarily imply independence, meaning that it may happen that $Cov(X,Y) = 0$ but X and Y are not independent.

3.3 Expectation of common random variables

3.3.1 Degenerate random variable

A random variable X is said to be degenerate if $X = c$ for some $c \in \mathbb{R}$: in this case, it is possible to state that $\mathbb{E}(X) = c$.
For proof, notice that, since $Im(X) = \{c\}$, the expectation will be given by:

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x\mathbb{P}(X = x) = 1 \cdot c = c$$

3.3.2 Bernoulli random variable

If $X \backsim Bernoulli(p)$, then $\mathbb{E}(X) = p$. For proof, notice that $Im(X) = \{0, 1\}$, meaning that the expectation will be given by:

$$\mathbb{E}(X) = \sum_{x \in Im(X)} x\mathbb{P}(X = x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

3.3.3 Binomial random variable

If $X \backsim Binomial(n, p)$, then $\mathbb{E}(X) = np$.
For proof, apply the definition of expectation:

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\mathbb{P}(X = k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!}p^k(1-p)^{n-k}$$

Rewrite $n! = n(n-1)!$ and apply a change of variable by letting $h = k - 1$ in order to rewrite the summation in the following way:

$$\mathbb{E}(X) = \sum_{h=0}^{n-1} \frac{n(n-1)!}{(n-(h+1))!h!}p^{h+1}(1-p)^{n-(h+1)} = np\sum_{h=0}^{n-1} \frac{(n-1)!}{(n-1-h)!h!}p^h(1-p)^{n-1-h}$$

Exploit the definition of the Newton binomial in order to conclude that:

$$\mathbb{E}(X) = np\sum_{h=0}^{n-1} \binom{n-1}{h}p^h(1-p)^{n-1-h} = np[p + (1-p)]^{n-1} = np$$

3.3.4 Geometric random variable

If $X \backsim Geometric(p)$, then $\mathbb{E}(X) = \frac{1}{p}$.
For proof, apply the definition of expectation:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Fix the summation to include $k = 0$ as well by noticing that it is possible to observe:

$$k(1-p)^{k-1} = -\frac{d((1-p)^k - 1)}{dp}$$

Knowing that $(1-p) \in (0, 1)$, it is possible to exchange summation and derivation to obtain:

$$\mathbb{E}(X) = p(-\frac{d(\sum_{k=0}^{\infty}(1-p)^k)}{dp}) = -p\frac{d(\frac{1}{1-(1-p)})}{dp} = -p\frac{d(\frac{1}{p})}{dp} = -p(-\frac{1}{p^2}) = \frac{1}{p}$$

3.3.5 Negative binomial random variable

If $X \backsim \bar{Bin}(k, p)$, then $\mathbb{E}(X) = \frac{k}{p}$.
For proof, a useful trick is to imagine X as the sum of identically distributed geometric random variables, meaning that:

$$X = \sum_{i=1}^{k} X_i, \text{ where } X_i \backsim Geometric(p).$$

At this point, knowing that $\mathbb{E}(X) = \frac{1}{p}$, it is possible to exploit the linearity of the expectation to conclude that:

$$\mathbb{E}(X) = \mathbb{E}(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} \mathbb{E}(X_i) = \sum_{i=1}^{k} \frac{1}{p} = \frac{k}{p}$$

3.3.6 Poisson random variable

If $X \backsim Poisson(\lambda)$, then $\mathbb{E}(X) = \lambda$.
For proof, apply the definition of expectation:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

N.B.: The summation will now start from $k = 1$ because the contribution from $k = 0$ is zero.
Apply a change of variable and let $k' = k-1$, allowing to rewrite the summation

in the following way:

$$\mathbb{E}(X) = e^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'+1}}{k'!} = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k'}}{k'!}$$

Most particularly, notice that the summation is actually a McLaurin expansion for $f(x) = e^x$, with $x = \lambda$, allowing to conclude that:

$$\sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} = e^{\lambda} \Rightarrow \mathbb{E}(X) = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

3.4 Variance of common random variables

3.4.1 Degenerate random variable

If $X = c$ is a degenerate random variable, then $\mathbb{V}(X) = 0$.
For proof, notice that a degenerate random variable is, by definition, a constant, meaning that it will never deviate from its mean, hence why $\mathbb{V}(X) = 0$.
Most particularly, notice that degenerate variables are the only random variables whose variance is exactly zero.

3.4.2 Bernoulli random variable

If $X \backsim Bernoulli(p)$, then $\mathbb{V}(X) = p(1-p)$.
For proof, apply the definition of variance:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = p - p^2 = p(1-p)$$

3.4.3 Binomial random variable

If $X \backsim Binomial(n, p)$, then $\mathbb{V}(X) = np(1-p)$.
For proof, a useful trick is to imagine X as the sum of identically distributed Bernoulli random variables, meaning that:

$$X = \sum_{i=1}^{n} X_i, \text{ where } X_i \backsim Bernoulli(p).$$

Most particularly, since these Bernoulli random variables are also independent, it is possible to exploit the fact that variance will be linear in order to conclude that:

$$\mathbb{V}(X) = \mathbb{V}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathbb{V}(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

3.4.4 Geometric random variable

If $X \backsim Geometric(p)$, then $\mathbb{V}(X) = \frac{1-p}{p^2}$.
For proof, apply the definition of variance:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

Since $\mathbb{E}(X)$ is already known, focus on finding $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \sum_{k=1}^{\infty} k^2 \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p$$

Let $k^2 = k(k-1) + k$ and split the summation into two simpler sums:

$$\mathbb{E}(X^2) = \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1} p + \sum_{k=1}^{\infty} k(1-p)^{k-1} p$$

Rearrange the summations to write the value as:

$$\mathbb{E}(X^2) = p\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2}(1-p) + \mathbb{E}(X) = p(1-p)\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} + \frac{1}{p}$$

Find the summation to include $k = 0$ as well in order to observe that:

$$k(k-1)(1-p)^{k-2} = \frac{d^2((1-p)^k - 1)}{dp^2}$$

Knowing that $p \in (0,1)$, it is possible to exchange summation and derivation to obtain that:

$$\mathbb{E}(X^2) = p(1-p)\cdot(\sum_{k=0}^{\infty} \frac{d^2((1-p)^k - 1)}{dp^2}) + \frac{1}{p} = p(1-p)\cdot(\frac{d^2(\sum_{k=0}^{\infty}(1-p)^k - 1)}{dp^2}) + \frac{1}{p}$$

By recognising a convergent geometric series, it is possible to simplify the result to:

$$\mathbb{E}(X^2) = p(1-p)\frac{d^2(\frac{1}{p})}{dp^2} + \frac{1}{p} = \frac{2p(1-p)}{p^3} + \frac{1}{p} = \frac{2-p}{p^2}$$

It is therefore possible to conclude that:

$$\mathbb{V}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

3.4.5 Negative binomial random variable

If $X \backsim \bar{Bin}(k,p)$, then $\mathbb{V}(X) = \frac{k(1-p)}{p^2}$.
For proof, a useful trick is to imagine X as the sum of identically distributed

geometric random variables, meaning that:

$$X = \sum_{i=1}^{k} X_i, \text{ where } X_i \backsim Geometric(p).$$

Most particularly, since the geometric random variables are also independent, it is possible to exploit the fact that variance will be linear in order toX = conclude that:

$$\mathbb{V}(X) = \mathbb{V}(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} \mathbb{V}(X_i) = \sum_{i=1}^{k} \frac{1-p}{p^2} = \frac{k(1-p)}{p^2}$$

3.4.6 Poisson random variable

If $X \backsim Poisson(\lambda)$, then $\mathbb{V}(X) = \lambda$.
For proof, apply the definition of variance:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

Since $\mathbb{E}(X)$ is already known, focus on finding $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2) = \sum_{k=0}^{\infty} k^2 \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!}$$

Let $k = 1 + k - 1$ and split the summation into two simpler sums:

$$\mathbb{E}(X^2) = e^{-\lambda}(\sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} + \sum_{k=0}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!})$$

Apply a change of variable to recover a McLaurin expansion for $f(x) = e^x$:

- For the first summation, let $k' = k - 1$:

  $$\sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} = \sum_{k'=0}^{\infty} \frac{\lambda^{k'+1}}{k'!} = \lambda \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!}$$

  Since the summation is now a McLaurin expansion for $f(x) = e^x$, with $x = \lambda$, it is possible to simplify it to the following result:

  $$\sum_{k=0}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{\lambda}$$

- For the second summation, let $k' = k - 2$:

  $$\sum_{k=0}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!} = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k-2)!} = \sum_{k'=0}^{\infty} \frac{\lambda^{k'+2}}{k'!} = \lambda^2 \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!}$$

Since the summation is now a McLaurin expansion for $f(x) = e^x$, with $x = \lambda$, it is possible to simplify it to the following result:

$$\sum_{k=0}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!} = \lambda^2 e^\lambda$$

The change of variable allows to rewrite the value as:

$$\mathbb{E}(X^2) = e^{-\lambda}(\lambda e^\lambda + \lambda^2 e^\lambda) = \lambda + \lambda^2$$

Therefore, it is possible to conclude that:

$$\mathbb{V}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$$

3.5 Poisson random variable as the limit of a Binomial random variable

$X_n \backsim Binomial(n, \frac{\lambda}{n})$ is approximated by $X \backsim Poisson(\lambda)$ as $n \to \infty$.
For proof, take the limit as $n \to \infty$ of the distribution function of a binomial random variable:

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \lim_{n\to\infty} \binom{n}{k} (\frac{\lambda}{n})^k (1 - \frac{\lambda}{n})^{n-k}$$

Consider the following approximation:

$$\lim_{n\to\infty} \binom{n}{k} \approx \frac{n^k}{k!}$$

Apply special limits:

$$\lim_{n\to\infty} (1 - \frac{\lambda}{n})^{n-k} = \lim_{n\to\infty} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} = \frac{e^{-\lambda}}{1} = e^{-\lambda} \text{ because } \lim_{n\to\infty} (1 + \frac{k}{n})^n = e^k$$

Therefore, it is possible to conclude that:

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \lim_{n\to\infty} \frac{n^k}{k!} \frac{\lambda^k}{n^k} e^{-\lambda} = \frac{e^{-\lambda}\lambda^k}{k!} = \mathbb{P}(X = k), \text{ with } X \backsim Poisson(\lambda)$$

3.5.1 Alternative proof for the expectation of a Poisson random variable

If $X \backsim Poisson(\lambda)$, then $\mathbb{E}(X) = \lambda$.
For proof, let $X_n \backsim Binomial(n, \frac{\lambda}{n})$ and consider its expectation for $n \to \infty$:

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} n\frac{\lambda}{n} = \lambda = \mathbb{E}(X), \text{ with } X \backsim Poisson(\lambda)$$

3.5.2 Alternative proof for the variance of a Poisson random variable

If $X \backsim Poisson(\lambda)$, then $\mathbb{V}(X) = \lambda$.
For proof, let $X_n \backsim Binomial(n, \frac{\lambda}{n})$ and consider its variance for $n \to \infty$:

$$\lim_{n\to\infty} \mathbb{V}(X_n) = \lim_{n\to\infty} n\frac{\lambda}{n}(1 - \frac{\lambda}{n}) = \lambda = \mathbb{V}(X), \text{ with } X \backsim Poisson(\lambda)$$

3.6 Sum of independent random variables

Let $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$ be two independent random variables and define $Z = X + Y$: $\mathbb{P}(Z = z) = \sum_{x\in S_x} \mathbb{P}(X = x)\mathbb{P}(Y = z - x)$.
For proof, given $Z : \Omega \to S_Z$, it is possible to state that:

$$\mathbb{P}(Z = z) = \mathbb{P}(X + Y = z) \ \forall \ z \in S_Z$$

Assume that the sample space Ω is partitioned with respect to the random variable X, meaning that it is possible to rewrite Ω as the following union of disjoint events:

$$\Omega = \bigcup_{x\in S_X} \{X = x\} \Rightarrow \mathbb{P}(X + Y = z, \Omega) = \mathbb{P}(\{X + Y = z\} \cap (\bigcup_{x\in S_X} \{X = x\}))$$

Therefore, it is possible to apply additivity in order to obtain:

$$\mathbb{P}(X + Y = z, \Omega) = \sum_{x\in S_X} \mathbb{P}(X + Y = z, X = x) = \sum_{x\in S_X} \mathbb{P}(X + Y = z, X = x)$$

However, since X and Y are independent random variables by assumption, it is possible to factorise the intersection and conclude that:

$$\mathbb{P}(Z = z) = \sum_{x\in S_X} \mathbb{P}(X + Y = z, X = x) = \sum_{x\in S_X} \mathbb{P}(X = x)\mathbb{P}(Y = z - x)$$

3.7 Markov's inequality

Given a non-negative random variable $X \geq 0$ and $\lambda > 0$, $\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}(X)}{\lambda}$.
For proof, consider the following computation:

$$\mathbb{P}(X \geq \lambda) = \sum_{x\in Im(X):\ x\geq\lambda} \mathbb{P}(X = x)$$

Most particularly, exploit the fact that $x \geq \lambda \Leftrightarrow \frac{x}{\lambda} \geq 1$ in order to get rid of the constraint of x, allowing to rewrite the previous statement as:

$$\mathbb{P}(X \geq \lambda) \leq \sum_{x\in Im(X)} \frac{x}{\lambda}\mathbb{P}(X = x)$$

However, since $\sum x\mathbb{P}(X = x) = \mathbb{E}(X)$ by definition, it is indeed possible to conclude that:

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}$$

3.8 Chebyshev's inequality

Given a random variable Y and $\lambda > 0$, $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \lambda) \leq \frac{\mathbb{V}(Y)}{\lambda^2}$.
For proof, define a non-negative random variable $X = |Y - \mathbb{E}(Y)|^2$, and apply Markov's inequality to state that:

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \lambda) \Rightarrow \mathbb{P}(X \geq \lambda^2) \leq \frac{\mathbb{E}(X)}{\lambda^2}$$

However, notice that, by definition of X, its expectation can actually be rewritten as:

$$\mathbb{E}(X) = \mathbb{E}(|Y - \mathbb{E}(Y)|^2) = \mathbb{V}(Y)$$

Therefore, it is indeed possible to conclude that:

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq \lambda) \leq \frac{\mathbb{V}(Y)}{\lambda^2}$$

3.9 The law of large numbers

Given a sequence $X_1, \ldots, X_n$ of identically distributed and independent random variables, each with finite expectation μ, if $S_n = \sum_{i=1}^n X_i$, then, $\forall\ \delta > 0$, it is possible to state that $\mathbb{P}(|\frac{S_n}{n} - \mu| < \delta) \to 1$ as $n \to \infty$.
For proof, exploit the linearity of expectation to state that:

$$\mathbb{E}(\frac{S_n}{n}) = \frac{1}{n}\mathbb{E}(\sum_{i=1}^{n} X_i) = \frac{n\mu}{n} = \mu \text{ by symmetry.}$$

Similarly, it is possible to exploit independence and symmetry of the random variables to state that:

$$\mathbb{V}(\frac{S_n}{n}) = \frac{1}{n^2}\mathbb{V}(\sum_{i=1}^{n} X_i) = \frac{n\mathbb{V}(X_1)}{n^2} = \frac{\mathbb{V}(X_1)}{n} \text{ by symmetry.}$$

Therefore, by applying Chebyshev's inequality on the random variable $\frac{S_n}{n}$, it is possible to obtain, for any $\delta > 0$, the following convergence:

$$\mathbb{P}(|\frac{S_n}{n} - \mu| \geq \delta) \leq \frac{\mathbb{V}(X_1)}{n\delta^2} \to 0 \text{ as } n \to \infty$$

By exploiting the complement of the previous event, it is therefore possible to conclude that:

$$\mathbb{P}(|\frac{S_n}{n} - \mu| < \delta) = 1 - \mathbb{P}(|\frac{S_n}{n} - \mu| \geq \delta) \to 1 \text{ as } n \to \infty$$

Chapter 4

Multinomial random variables

4.1 Recovering the marginal distribution of one variable

If $(X_1, \dots, X_k) \backsim \textit{Multinomial}(n, p_1, \dots, p_k)$*, with* $n = \sum_{i=1}^{k} n_i$*, it is possible to recover the marginal distribution of* X_i *as* $\mathbb{P}(X_i = n_i) = \binom{n}{n_i} p_i^{n_i}(1 - p_i)^{n-n_i}$*.*
N.B.: For simplicity, the statement is proved for $k = 3$.
For proof, start by considering the joint multinomial distribution of X_1, X_2, X_3:

$$\mathbb{P}(X_1 = n_1, X_2 = n_2, X_3 = n_3) = \binom{n}{n_1\ n_2\ n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3} = \frac{n!}{n_1! \cdot n_2! \cdot n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

Since $X_1 = n_1$ by assumption, fix $X_2 = n_2$ as well and, by definition, let $X_3 = n - (n_1 + n_2)$ in order to recover the marginal distribution of X_1 through the law of total probability:

$$\mathbb{P}(X_1 = n_1) = \sum_{n_2=0}^{n-n_1} \mathbb{P}(X_1 = n_1, X_2 = n_2, X_3 = n-(n_1+n_2)) = \sum_{n_2=0}^{n-n_1} \frac{n!}{n_1! \cdot n_2! \cdot (n - (n_1 + n_2))!} p_1^{n_1} p_2^{n_2} p_3^{n-(n_1+n_2)}$$

Take out constant terms and manipulate the summation in order to reconduct the multinomial distribution to a binomial one:

$$\mathbb{P}(X_1 = n_1) = \frac{n! \cdot p_1^{n_1}}{n_1!} \sum_{n_2=0}^{n-n_1} \frac{1}{n_2! \cdot (n - n_1 - n_2)!} p_2^{n_2} p_3^{n-n_1-n_2} \cdot \frac{(n - n_1)!}{(n - n_1)!}$$

Start to reconstruct the binomial coefficients and exploit the definition of Newton binomial in order to simplify:

$$\mathbb{P}(X_1 = n_1) = \frac{n! \cdot p_1^{n_1}}{n_1!(n - n_1)!} \sum_{n_2=0}^{n-n_1} \frac{(n - n_1)!}{n_2! \cdot (n - n_1 - n_2)!} p_2^{n_2} p_3^{n-n_1-n_2} = \binom{n}{n_1} p_1^{n_1} (p_2+p_3)^{n-n_1}$$

Most particularly, if $p_2 + p_3 = 1 - p_1$, this probability can ultimately be written as:

$$\mathbb{P}(X_1 = n_1) = \binom{n}{n_1} p_1^{n_1} (1 - p_1)^{n-n_1}$$

4.2 Conditional multinomial distribution

If $(X_1, \ldots, X_k) \backsim Multinomial(n, p_1, \ldots, p_k)$*, with* $X_2 = n_2$ *being a known event, then* $\mathbb{P}(X_1 = n_1|X_2 = n_2) = \binom{n-n_2}{n_1}(\frac{p_1}{p_1+p_3+\cdots+p_k})^{n_1}(\frac{p_3+\cdots+p_k}{p_1+p_3+\cdots+p_k})^{n-n_2-n_1}$.
N.B.: For simplicity, the statement is proved for $k = 3$**.**
For proof, start by considering the joint multinomial distribution of X_1, X_2, X_3:

$$\mathbb{P}(X_1 = n_1, X_2 = n_2, X_3 = n_3) = \binom{n}{n_1\ n_2\ n_3} p_1^{n_1} p_2^{n_2} p_3^{n_3} = \frac{n!}{n_1! \cdot n_2! \cdot n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

Since $X_1 = n_1$ by assumption and $X_2 = n_2$ is given, let $X_3 = n - (n_1 + n_2)$ and apply the definition of conditional probability:

$$\mathbb{P}(X_1 = n_1|X_2 = n_2) = \frac{\mathbb{P}(X_1 = n_1, X_2 = n_2)}{\mathbb{P}(X_2 = n_2)}$$

By applying the joint and marginal distributions, this value will become:

$$\mathbb{P}(X_1 = n_1|X_2 = n_2) = \frac{\frac{n!}{n_1! \cdot n_2! \cdot (n-(n_1+n_2))!} p_1^{n_1} p_2^{n_2} p_3^{n-(n_1+n_2)}}{\frac{n!}{n_2!(n-n_2)!} p_2^{n_2} (p_1 + p_3)^{n-n_2}}$$

Start to simplify the equation in order to recover a binomial distribution:

$$\mathbb{P}(X_1 = n_1|X_2 = n_2) = \frac{(n - n_2)!}{n_1!(n - n_2 - n_1)!} \frac{p_1^{n_1} p_3^{n-n_1-n_2}}{(p_1 + p_3)^{n-n_2}}$$

The result therefore simplifies to:

$$\mathbb{P}(X_1 = n_1|X_2 = n_2) = \binom{n - n_2}{n_1} (\frac{p_1}{p_1 + p_3})^{n_1} (\frac{p_3}{p_1 + p_3})^{n-n_2-n_1}$$

Chapter 5

Continuous probability

5.1 Legitness of continuous random variables

5.1.1 Uniform random variable

If $X \backsim Uniform(a, b)$*, then* $f_X(x) = \frac{1}{b-a}$ *defines a legit continuous probability distribution.*

For proof, start by considering that a uniform random variable is defined to assign a constant value $c \in \mathbb{R} \ \forall \ x \in [a, b]$, meaning that its probability density function should be of the type:

$$f_X(x) = \begin{cases} c & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Keep in mind, however, that a probability density function is legit if and only if its cumulative distribution function satisfies the following statement:

$$F_X(\mathbb{R}) = \int_{-\infty}^{\infty} f_X(x) \ dx = 1$$

Therefore, it must hold that:

$$\int_{-\infty}^{\infty} f_X(x) \ dx = \int_{a}^{b} c \ dx = [cx]_a^b = c(b-a) = 1 \Leftrightarrow c = \frac{1}{b-a}$$

Therefore, the probability density function of a uniform distribution will be given by:

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

Notice that this is indeed a legit distribution because it is possible to recover that:

$$\int_{-\infty}^{\infty} f_X(x)\ dx = \int_a^b \frac{1}{b-a}\ dx = \frac{1}{b-a}[x]_a^b = \frac{b-a}{b-a} = 1$$

5.2 Expectation of a continuous random variable

5.2.1 Uniform random variable

If $X \backsim Uniform(a,b)$, then $\mathbb{E}(X) = \frac{a+b}{2}$.
For proof, start by considering the probability density function of $X \backsim Uniform(a,b)$:

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

At this point, apply the definition of expectation:

$$\mathbb{E}(X) = \int_a^b x f_X(x)\ dx = \int_a^b \frac{x}{b-a}\ dx = \frac{1}{b-a}[\frac{x^2}{2}]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

5.2.2 Gaussian random variable

If $X \backsim \mathcal{N}(\mu,\sigma^2)$, then $\mathbb{E}(X) = \mu$.
For proof, apply the definition of expectation:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx = \mu \text{ by symmetry of the Gaussian integral.}$$

5.3 Variance of a continuous random variable

5.3.1 Uniform random variable

If $X \backsim Uniform(a,b)$, then $\mathbb{V}(X) = \frac{(b-a)^2}{12}$.
For proof, start by considering the probability density function of $X \backsim Uniform(a,b)$:

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

At this point, apply the definition of variance:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \int_a^b x^2 f_X(x)\ dx - (\int_a^b x f_X(x)\ dx)^2$$

Therefore, it is possible to conclude that:

$$\mathbb{V}(X) = \int_a^b \frac{x^2}{b-a}\ dx - (\int_a^b \frac{x}{b-a})^2 = [\frac{x^3}{3(b-a)}]_a^b - ([\frac{x^2}{2(b-a)}]_a^b)^2 = \frac{(b-a)^2}{12}$$

5.3.2 Gaussian random variable

If $X \backsim \mathcal{N}(\mu, \sigma^2)$*, then* $\mathbb{V}(X) = \sigma^2$*.*
For proof, evaluate the variance of X through the centered variable:

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} (X - \mathbb{E}(X))^2 f_X(x)\ dx = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx$$

Apply a change of variable and let $u = \frac{x-\mu}{\sigma}$, meaning that $\frac{du}{dx} = \frac{1}{\sigma}$, allowing to rewrite the integral as:

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} \sigma^2 u^2 \frac{1}{2\pi} e^{-\frac{u^2}{2}}\ du = \sigma^2 \int_{-\infty}^{\infty} u \cdot \frac{ue^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\ du$$

At this point, it is possible to solve the integral by applying integration by parts:

$$\mathbb{V}(X) = \sigma^2 ([\frac{-ue^{-\frac{u^2}{2}}}{\sqrt{2\pi}}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\ du)$$

Notice, however, that the first integral is actually the expectation of $Z \backsim \mathcal{N}(0, 1)$, meaning that it will be equal to 0, whereas the second integral is a Gaussian integral, which, by definition, is equal to 1.
Therefore, it is possible to conclude that, indeed, $\mathbb{V}(X) = \sigma^2$.