Chapter 2

Random Variables

2.1 Probability Distribution and Expectation

2.1.1 Random Variables and their Distributions

The number of heads in a sequence of 10000 coin tosses, the number of days it takes until the next rain and the size of a genealogical tree are random numbers. All are functions of the outcome of a random experiment (performed either by man or by nature) and taking discrete values, that is, values in a countable set. These values are integers in the above examples, but they can be more complex mathematical objects, such as graphs for instance. This chapter gives the elementary rules for computing expectations, a list of famous discrete random variables or vectors (binomial, geometric, Poisson and multinomial), and the elementary theory of conditional expectation.

Definition 2.1.1 *Let E be a countable set. A function $X : \Omega \to E$ such that for all $x \in E$*

$$\{\omega; X(\omega) = x\} \in \mathcal{F}$$

is called a discrete random variable.

Since E is a countable set, it can always be identified with $\mathbb{N}$ or $\overline{\mathbb{N}}$, and therefore we shall often assume that either $E = \mathbb{N}$ or $\overline{\mathbb{N}}$.

Being in $\mathcal{F}$, the event $\{X = x\}$ can be assigned a probability.

Remark 2.1.2 Calling a random variable a *random number* is an innocuous habit as long as one is aware that it is *not the function* X that is random, but the outcome ω. This in turn makes the *number* $X(\omega)$ random.

EXAMPLE 2.1.3: TOSSING A DIE, TAKE 3. The sample space is the set $\Omega = \{1, 2, 3, 4, 5, 6\}$. Take for X the identity: $X(\omega) = \omega$. In that sense X is the random number obtained by tossing a die.

P. Brémaud, *Discrete Probability Models and Methods*,
Probability Theory and Stochastic Modelling 78,
DOI 10.1007/978-3-319-43476-6_2

EXAMPLE 2.1.4: HEADS AND TAILS, TAKE 4. The sample space Ω is the collection of all sequences $\omega = \{x_n\}_{n\geq 1}$, where $x_n = 1$ or 0. Define a random variable X_n by $X_n(\omega) = x_n$. It is the random number obtained at the n-th toss. It is indeed a random variable since for all $a_n \in \{0, 1\}$, $\{\omega\,;\, X_n(\omega) = a_n\} = \{\omega\,;\, x_n = a_n\} \in \mathcal{F}$, by definition of $\mathcal{F}$.

The following are elementary remarks.

Theorem 2.1.5 *Let E and F be countable sets. Let X be a random variable with values in E, and let $f : E \to F$ be an arbitrary function. Then $Y := f(X)$ is a random variable.*

Proof. Let $y \in F$. The set $\{\omega;\, Y(\omega) = y\}$ is in $\mathcal{F}$ since it is a countable union of sets in $\mathcal{F}$, namely:

$$\{Y = y\} = \sum_{x\in E;\, f(x)=y} \{X = x\}\,.$$

□

Theorem 2.1.6 *Let E_1 and E_2 be countable sets. Let X_1 and X_2 be random variable with values in E_1 and E_2 respectively. Then $Y := (X_1, X_2)$ is a random variable with values in $E = E_1 \times E_2$.*

Proof. Let $x = (x_1, x_2) \in E$. The set $\{\omega;\, X(\omega) = x\}$ is in $\mathcal{F}$ since it is the intersection of sets in $\mathcal{F}$: $\{X = x\} = \{X_1 = x_1\} \cap \{X_2 = x_2\}$. □

Definition 2.1.7 *From the probabilistic point of view, a discrete random variable X is described by its probability distribution function (or distribution, for short) $\{\pi(x)\}_{x\in E}$, where $\pi(x) := P(X = x)$.*

EXAMPLE 2.1.8: THE UNIFORM DISTRIBUTION. Let $\mathcal{X}$ be a finite set. The random variable with values in this set and having the distribution

$$P(X = x) = \frac{1}{|\mathcal{X}|} \text{ for all } x \in \mathcal{X}$$

is said to be uniformly distributed (or to have the uniform distribution) on $\mathcal{X}$.

EXAMPLE 2.1.9: IS THIS NUMBER THE LARGER ONE? Let a and b be two numbers in $\{1, 2, \ldots, 10, 000\}$. Nothing is known about these numbers, except that they are not equal, say $a > b$. Only one of these numbers is shown to you, secretely chosen at random and equiprobably. Call X this random number. Is there a good strategy for guessing if the number shown to you is the largest of the two? Of course, we would like to have a probability of success strictly larger than $\frac{1}{2}$.

Perhaps surprisingly, there is such a strategy, that we now describe. Select at random uniformly on $\{1, 2, \dots, 10,000\}$ a number Y. If $X \geq Y$, say that X is the largest $(= a)$, otherwise say that it is the smallest.

Let us compute the probability P_E of a wrong guess. An error occurs when either (i) $X \geq Y$ and $X = b$, or (ii) $X < Y$ and $X = a$. These events are exclusive of one another, and therefore

$$\begin{aligned} P_E &= P(X \geq Y, X = b) + P(X < Y, X = a) \\ &= P(b \geq Y, X = b) + P(a < Y, X = a) \\ &= P(b \geq Y)P(X = b) + P(a < Y)P(X = a) \\ &= P(b \geq Y)\frac{1}{2} + P(a < Y)\frac{1}{2} = \frac{1}{2}(P(b \geq Y) + P(a < Y)) \\ &= \frac{1}{2}(1 - P(Y \in [b+1, a]) = \frac{1}{2}\left(1 - \frac{a-b}{10,000}\right) < \frac{1}{2}\,. \end{aligned}$$

EXAMPLE 2.1.10: HEADS AND TAILS, TAKE 5. The number of occurrences of heads in n tosses is $S_n = X_1 + \cdots + X_n$. This random variable is the fortune at time n of a gambler systematically betting on heads. It takes the integer values from 0 to n. We have

$$P(S_n = k) = \tfrac{1}{2^n}\tbinom{n}{k}\,.$$

Proof. The event $\{S_n = k\}$ is "k among $X_1, \dots, X_n$ are equal to 1." There are $\binom{n}{k}$ distinct ways of assigning k values 1 and $n - k$ values 0 to $X_1, \dots, X_n$, and all have the same probability 2^{-n}. □

One sometimes needs to prove that a random variable X taking its values in $\overline{\mathbb{N}}$ (the value ∞ is *a priori* possible) is in fact almost surely finite, that is, one must prove that $P(X = \infty) = 0$ or, equivalently, $P(X < \infty) = 1$. Since $\{X < \infty\} = \sum_{n=0}^{\infty}\{X = n\}$, we have $P(X < \infty) = \sum_{n=0}^{\infty} P(X = n)$.

Remark 2.1.11 We seize this opportunity to recall that in an expression such as $\sum_{n=0}^{\infty}$, the sum is over $\mathbb{N}$ and does not include ∞ as the notation seems to suggest. A less ambiguous notation would be $\sum_{n\in\mathbb{N}}$. In case we want to sum over all integers plus ∞, we shall *always* use the notation $\sum_{n\in\overline{\mathbb{N}}}$.

The following result is highlighted as a theorem for the purpose of future reference:

Theorem 2.1.12 *Let X be an integer-valued random variable (in particular, the probability that $X = \infty$ is null). Then*

$$\lim_{n\uparrow\infty} P(X>n) = 0\,.$$

Proof. This follows by monotone sequential continuity since the sequence $\{X > n\}$, $n \geq 0$, is non-increasing and $\cap_{n\geq 0}\{X > n\} = \varnothing$ since X takes only finite values. □

Almost Surely, take 2

An expression like "$X = Y$ P-almost surely" means that $P(\{\omega \in \Omega;\, X(\omega) = Y(\omega)\}) = 1$. One interprets similarly expressions such as "$f(X) = 0$ P-almost surely" and so on.

2.1.2 Independent Random Variables

Definition 2.1.13 *Two discrete random variables X and Y are called independent if for all $i, j \in E$,*

$$P(X=i, Y=j) = P(X=i)P(Y=j)\,. \tag{2.1}$$

The extension of the definition to a finite number of random variables is natural:

Definition 2.1.14 *The discrete random variables $X_1, \dots, X_k$ taking their values in $E_1, \dots, E_k$ respectively are said to be independent if for all $i_1 \in E_1, \dots, i_k \in E_k$,*

$$P(X_1 = i_1, \dots, X_k = i_k) = P(X_1 = i_1) \cdots P(X_k = i_k)\,. \tag{2.2}$$

Theorem 2.1.15 *Let $X_1, \dots, X_k$ be as in Definition 2.1.14. Then, for any $g_i : E_i \to \mathbb{R}$ $(1 \leq i \leq n)$, the random variables $g_i(X_i)$ $(1 \leq i \leq n)$ are independent.*

Proof. We do the proof in the case $n = 2$:

$$\begin{aligned}
P(g_1(X_1) = j_1, g_2(X_2) = j_2) &= \sum_{i_1; g_1(i_1)=j_1} \; \sum_{i_2; g_1(i_2)=j_2} P(X_1 = i_1, X_2 = i_2) \\
&= \sum_{i_1; g_1(i_1)=j_1} \; \sum_{i_2; g_1(i_2)=j_2} P(X_1 = i_1)P(X_2 = i_2) \\
&= \left(\sum_{i_1; g_1(i_1)=j_1} P(X_1 = i_1) \right) \left(\sum_{i_2; g_1(i_2)=j_2} P(X_2 = i_2) \right) \\
&= P(g_1(X_1) = j_1) P(g_2(X_2) = j_2)\,.
\end{aligned}$$

□

Definition 2.1.16 *A sequence $\{X_n\}_{n\geq 1}$ of discrete random variables indexed by the set of positive integers and taking their values in the sets $\{E_n\}_{n\geq 1}$ respectively is called independent if for all $n \geq 2$, the random variables $X_1, \dots, X_n$ are independent. If in addition $E_n \equiv E$ for all $n \geq 1$ and the distribution of X_n does not depend on n, the sequence $\{X_n\}_{n\geq 1}$ is said to be* IID *(independent and identically distributed).*

EXAMPLE 2.1.17: HEADS AND TAILS, TAKE 6. We show that the sequence $\{X_n\}_{n\geq 1}$ is IID. Therefore, we have a model for *independent* tosses of an *unbiased* coin.

Proof. Event $\{X_k = a_k\}$ is the direct sum of events $\{X_1 = a_1, \dots, X_{k-1} = a_{k-1}, X_k = a_k\}$ for all possible values of $(a_1, \dots, a_{k-1})$. Since there are 2^{k-1} such values and each one has probability 2^{-k}, we have $P(X_k = a_k) = 2^{k-1}2^{-k}$, that is,

$$P(X_k = 1) = P(X_k = 0) = \frac{1}{2}.$$

Therefore,

$$P(X_1 = a_1, \dots, X_k = a_k) = P(X_1 = a_1) \cdots P(X_k = a_k)$$

for all $a_1, \dots, a_k \in \{0, 1\}$, from which it follows by definition that $X_1, \dots, X_k$ are independent random variables, and more generally that $\{X_n\}_{n\geq 1}$ is a family of independent random variables. □

Definition 2.1.18 *Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ be sequences of discrete random variables indexed by the positive integers and taking their values in the sets $\{E_n\}_{n\geq 1}$ and $\{F_n\}_{n\geq 1}$ respectively. They are said to be independent if for any finite collection of random variables $X_{i_1}, \dots, X_{i_r}$ and $Y_{j_1}, \dots, Y_{i_s}$ extracted from their respective sequences, the discrete random variables $(X_{i_1}, \dots, X_{i_r})$ and $(Y_{j_1}, \dots, Y_{i_s})$ are independent.*

(This means that

$$\begin{aligned} &P\left((\cap_{\ell=1}^{r}\{X_{i_\ell} = a_\ell\}) \cap (\cap_{m=1}^{s}\{Y_{j_m} = b_m\})\right) \\ &= P\left(\cap_{\ell=1}^{r}\{X_{i_\ell} = a_\ell\}\right) P\left(\cap_{m=1}^{s}\{Y_{j_m} = b_m\}\right) \end{aligned} \tag{2.3}$$

for all $a_1 \in E_1, \dots, a_r \in E_r$, $b_1 \in F_1, \dots, b_s \in F_s$.)

The notion of conditional independence for events (Definition 1.3.14) extends naturally to discrete random variables.

Definition 2.1.19 *Let X, Y, Z be random variables taking their values in the denumerable sets E, F, G, respectively. One says that X and Y are conditionally independent given Z if for all x, y, z in E, F, G, respectively, events $\{X = x\}$ and $\{Y = y\}$ are conditionally independent given $\{Z = z\}$.*

2.1.3 Expectation

Definition 2.1.20 *Let X be a discrete random variable taking its values in the countable set E and let $g : E \to \mathbb{R}$ be a function that is either non-negative or such that*

$$\textstyle\sum_{x\in E} |g(x)|P(X = x) < \infty\,. \tag{2.4}$$

Then one defines $E[g(X)]$, the expectation of $g(X)$, by the formula

$$E[g(X)] = \textstyle\sum_{x\in E} g(x)P(X = x)\,. \tag{2.5}$$

If the summability condition (2.4) is satisfied, we say that the random variable $g(X)$ is integrable, and in this case the expectation $E[g(X)]$ is a *finite* number. If it is only assumed that g is non-negative, the expectation may well be infinite.

Example 2.1.21: Heads and tails, take 7. Consider the random variable $S_n = X_1 + \cdots + X_n$ with values in $\{0, 1, \dots, n\}$. Its expectation is $E[S_n] = n/2$. In fact,

$$\begin{aligned} E[S_n] &= \sum_{k=0}^{n} kP(S_n = k) = \frac{1}{2^n}\sum_{k=1}^{n} k\frac{n!}{k!(n-k)!} \\ &= \frac{n}{2^n}\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \\ &= \frac{n}{2^n}\sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} = \frac{n}{2^n}2^{n-1}. \end{aligned}$$

Example 2.1.22: Finite random variables with infinite expectations. It is important to realize that a discrete random variable taking *finite values* may have an *infinite expectation.* The canonical example is the random variable X with values in $E = \mathbb{N}_+$ and such that

$$P(X = n) = \frac{1}{cn^2} \qquad (n \in \mathbb{N}_+)$$

where the constant c is chosen such that X actually takes its values in $\mathbb{N}$:

$$P(X < \infty) = \sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty} \frac{1}{cn^2} = 1$$

(therefore $c = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$). In fact, the expectation of X is

$$E[X] = \sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} n\frac{1}{cn^2} = \sum_{n=1}^{\infty} \frac{1}{cn} = \infty.$$

Remark 2.1.23 Although the above example is artificial, there are many natural occurences of the phenomenon. Consider for instance Example 2.1.21, and let T be the first integer n such that $2S_n - n = 0$. Then, as it turns out, and as we shall prove in Subsection 8.1.1 that T is a *finite* random variable with *infinite expectation*. Note that the quantity $2S_n - n$ is the fortune at time n of a gambler systematically betting one *euro* on heads.

The telescope formula below gives an alternative way of computing the expectation of an integer-valued random variable.

Theorem 2.1.24 *For a random variable X taking its values in $\mathbb{N}$,*

$$E[X] = \textstyle\sum_{n=1}^{\infty} P(X \geq n)\,.$$

Proof.

$$\begin{aligned} E[X] &= P(X=1) + 2P(X=2) + 3P(X=3) + \dots \\ &= P(X=1) \ \ + P(X=2) + P(X=3) + \dots \\ &\qquad\qquad\quad + P(X=2) + P(X=3) + \dots \\ &\qquad\qquad\qquad\qquad\qquad + P(X=3) + \dots \end{aligned}$$

□

We now list a few elementary properties of expectation.

Theorem 2.1.25 *Let A be some event. The expectation of the indicator random variable $X = 1_A$ is*

$$E[1_A] = P(A)\,. \tag{2.6}$$

Proof. $X = 1_A$ takes the value 1 with probability $P(X=1) = P(A)$ and the value 0 with probability $P(X=0) = P(\overline{A}) = 1 - P(A)$. Therefore,

$$E[X] = 0 \times P(X=0) + 1 \times P(X=1) = P(X=1) = P(A).$$

□

Theorem 2.1.26 *Let g_1 and g_2 be functions from E to $\overline{\mathbb{R}}$ such that $g_1(X)$ and $g_2(X)$ are integrable (resp., non-negative), and let $\lambda_1, \lambda_2 \in \mathbb{R}$ (resp., $\in \mathbb{R}_+$). Expectation is linear, that is,*

$$E[\lambda_1 g_1(X) + \lambda_2 g_2(X)] = \lambda_1 E[g_1(X)] + \lambda_2 E[g_2(X)]\,. \tag{2.7}$$

Also, expectation is monotone, in the sense that if $g_1(x) \leq g_2(x)$ for all x such that $P(X = x) > 0$ (in other words, $g_1(X) \leq g_2(X)$ almost surely)

$$E[g_1(X)] \leq E[g_2(X)]\,. \tag{2.8}$$

Also, we have the triangle inequality

$$|E[g(X)]| \leq E[|g(X)|]\,. \tag{2.9}$$

Proof. These properties follow from the corresponding properties of series. □

Theorem 2.1.27 *Let X be a random variable with values in E and let $g : E \to \overline{\mathbb{R}}_+$ be a non-negative function.*

(a) If $E\left[g(X)\right] = 0$, then $g(X) = 0$ P-almost surely.

(b) If $E\left[g(X)\right] < \infty$, then $g(X) < \infty$ P-almost surely.

Proof. (a) Condition $E\left[g(X)\right] = 0$ reads $\sum_{x\in E} g(x)P(X = x) = 0$. In particular $P(X = x) = 0$ whenever $g(x) > 0$. Therefore

$$P(g(X) > 0) = \sum_{x\in E;\, g(x)>0} P(X = x) = 0$$

or, equivalently, $P(g(X) = 0) = 1$.

(b) Condition $E\left[g(X)\right] < \infty$ reads $\sum_{x\in E} g(x)P(X = x) < \infty$. In particular $P(X = x) = 0$ whenever $g(x) = \infty$. Therefore

$$P(g(X) = \infty) = \sum_{x\in E;\, g(x)=\infty} P(X = x) = 0$$

or, equivalently, $P(g(X) < \infty) = 1$. □

Product Formula for Expectations

Theorem 2.1.28 *Let Y and Z be two independent random variables with values in the (denumerable) sets F and G respectively, and let $v : F \to \overline{\mathbb{R}}$, $w : G \to \overline{\mathbb{R}}$ be functions that are either non-negative, or such that $v(Y)$ and $w(Z)$ are both integrable. Then*

$$E[v(Y)w(Z)] = E[v(Y)]E[w(Z)]\,.$$

Proof. Consider the discrete random variable X with values in $E = F \times G$ defined by $X = (Y, Z)$, and consider the function $g : E \to \overline{\mathbb{R}}$ defined by $g(x) = v(y)w(z)$ where $x = (y, z)$. Under the above stated conditions, we have

$$\begin{aligned}
E[v(Y)w(Z)] &= E[g(X)] = \sum_{x\in E} g(x)P(X = x) \\
&= \sum_{y\in F}\sum_{z\in F} v(y)w(z)P(Y = y, Z = z) \\
&= \sum_{y\in F}\sum_{z\in F} v(y)w(z)P(Y = y)P(Z = z) \\
&= \left(\sum_{y\in F} v(y)P(Y = y)\right)\left(\sum_{z\in F} w(z)P(Z = z)\right) \\
&= E[v(Y)]E[w(Z)].
\end{aligned}$$

□

Mean, Variance and Covariance

Definition 2.1.29 *Let X be an integrable random variable. In this case, we define its mean as the (finite) number*

$$\mu = E[X] .$$

Let X be a square-integrable random variable. We then define its variance σ^2 by

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

(In the case of integer-valued random variables, the mean and variance, when they are well-defined, are therefore given by the following sums:

$$\mu = \sum_{n=0}^{+\infty} nP(X = n) \qquad \sigma^2 = \sum_{n=0}^{+\infty} (n - \mu)^2 P(X = n) .)$$

The variance is also denoted by $\text{Var}\ (X)$. From the linearity of expectation, it follows that $E[(X - m)^2] = E[X^2] - 2mE[X] + m^2$, that is,

$$\text{Var}\ (X) = E[X^2] - m^2 .$$

The mean is the "center of inertia" of a random variable. More precisely,

Theorem 2.1.30 *Let X be a real integrable random variable with mean m and finite variance σ^2. Then, for all $a \in \mathbb{R}$, $a \neq \mu$,*

$$E[(X - a)^2] > E[(X - \mu)^2] = \sigma^2 .$$

Proof.

$$\begin{aligned} E\left[(X - a)^2\right] &= E\left[((X - \mu) + (\mu - a))^2\right] \\ &= E\left[(X - \mu)^2\right] + (\mu - a)^2 + 2(\mu - a)E\left[(X - \mu)\right] \\ &= E\left[(X - \mu)^2\right] + (\mu - a)^2 > E\left[(X - \mu)^2\right] \end{aligned}$$

whenever $a \neq \mu$. □

The following consequence of the product rule is extremely important. It says that for *independent* random variables, variances add up.

Theorem 2.1.31 *Let $X_1, \dots, X_n$ be independent square-integrable random variables. Then*

$$\sigma^2_{X_1 + \cdots + X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n} .$$

Proof. Let $\mu_1, \dots, \mu_n$ be the respective means of $X_1, \dots, X_n$. The mean of the sum $X := X_1 + \cdots + X_n$ is $\mu := \mu_1 + \cdots + \mu_n$. If $i \neq k$, we have, by the product formula for expectations,

$$E\left[(X_i - \mu_i)(X_k - \mu_k)\right] = E\left[(X_i - \mu_i)\right] E\left[(X_k - \mu_k)\right] = 0.$$

Therefore

$$\begin{aligned} \text{Var}\,(X) =& E\left[(X-\mu)^2\right] = E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right] \\ =& E\left[\sum_{i=1}^{n}\sum_{k=1}^{n}(X_i - \mu_i)(X_k - \mu_k)\right] \\ =& \sum_{i=1}^{n}\sum_{k=1}^{n} E\left[(X_i - \mu_i)(X_k - \mu_k)\right] \\ =& \sum_{i=1}^{n} E\left[(X_i - \mu_i)^2\right] = \sum_{i=1}^{n} \text{ Var}\,(X_i). \end{aligned}$$

□

Note that means always add up, even when the random variables are not independent.

Let X be an integrable random variable. Then, clearly, for any $a \in \mathbb{R}$, aX is integrable and its variance is given by the formula

$$\text{Var}\,(aX) = a^2 \text{ Var}\,(X)\,.$$

EXAMPLE 2.1.32: VARIANCE OF THE EMPIRICAL MEAN. From this remark and Theorem 2.1.31, it immediately follows that if $X_1, \dots, X_n$ are independent and identically distributed *integrable* random variables with values in $\mathbb{N}$ with common variance σ^2, then

$$\text{Var}\,\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}\,.$$

2.1.4 Famous Distributions

A random variable X taking its values in $\{0, 1\}$ with distribution given by

$$P(X = 1) = p,$$

where $p \in (0, 1)$, is called a Bernoulli random variable with parameter p. This is denoted

$$X \sim Bern(p)\,.$$

Consider the following heads and tails framework which consists of an IID sequence $\{X_n\}_{n\geq 1}$ of Bernoulli variables with parameter p. It is called a Bernoulli sequence with parameter p.

Since $P(X_j = a_j) = p$ or $1-p$ depending on whether $a_i = 1$ or 0, and since there are exactly $h(a) := \sum_{j=1}^{k} a_j$ coordinates of $a = (a_1, \dots, a_k)$ equal to 1,

$$P(X_1 = a_1, \dots, X_k = a_k) = p^{h(a)} q^{k-h(a)} \, ,$$

where $q := 1 - p$. (The integer $h(a)$ is called the Hamming weight of the binary vector a.) Comparing with Examples 1.1.3 and 1.2.3, we see that we have a probabilistic model of a game of heads and tails, with a biased coin when $p \neq \frac{1}{2}$.

The heads and tails framework gives rise to two famous discrete random variables: the binomial random variable, and the geometric random variable.

The Binomial Distribution

Definition 2.1.33 *A random variable X taking its values in the set $E = \{0, 1, \dots, n\}$ and with the distribution*

$$P(X = i) = \tbinom{n}{i} p^i (1-p)^{n-i}$$

is called a binomial random variable of size n and parameter $p \in (0, 1)$.

This is denoted

$$X \sim \mathcal{B}(n, p) \, .$$

EXAMPLE 2.1.34: We place ourselves in the heads and tails framework. Define

$$S_n = X_1 + \dots + X_n \, .$$

This random variable takes the values $0, 1, \dots, n$. To obtain $S_n = i$ where $0 \leq i \leq n$, one must have $X_1 = a_1, \dots, X_n = a_n$ with $\sum_{j=1}^{n} a_j = i$. There are $\binom{n}{i}$ distinct ways of having this, each one occuring with probability $p^i(1-p)^{n-i}$. Therefore, for $0 \leq i \leq n$,

$$P(S_n = i) = \tbinom{n}{i} p^i (1-p)^{n-i}.$$

Theorem 2.1.35 *The mean and the variance of a binomial random variable X of size n and parameter p are given by*

$$E[X] = np \, ,$$

$$\text{Var}\,(X) = np(1-p) \, .$$

Proof. This can be proven by a direct computation. Later on, in the Exercises section, you will prove this using generating functions. Another approach is to start from the random variable S_n of Example 2.1.34. This is a binomial random variable. We have

$$E\left[S_n\right] = \sum_{i=1}^{n} E\left[X_i\right] = nE\left[X_1\right]$$

and, since the X_i's are IID,

$$V(S_n) = \sum_{i=1}^{n} V(X_i) = nV(X_1)\,.$$

Now,

$$E\left[X_1\right] = 0 \times P(X_1 = 0) + 1 \times P(X_1 = 1) = P(X_1 = 1) = p$$

and, since $X_1^2 = X_1$,

$$E\left[X_1^2\right] = E\left[X_1\right] = p\,.$$

Therefore

$$V(X_1) = E\left[X_1^2\right] - E\left[X_1\right]^2 = p - p^2 = p(1-p)\,.$$

□

The following inequalities concerning the binomial coefficients are useful:

Theorem 2.1.36 *Let $p \in (0,1)$ and $H_2(p) := -p\log_2 p - q\log_2 q$, where $q := 1-p$. Then for $0 < p \leq \frac{1}{2}$,*

$$\binom{n}{\lfloor np \rfloor} \leq 2^{nH_2(p)}. \tag{2.10}$$

For $\frac{1}{2} \leq p < 1$,

$$\binom{n}{\lceil np \rceil} \leq 2^{nH_2(p)}\,. \tag{$\star$}$$

For $\frac{1}{2} \leq p < 1$,

$$\frac{2^{nH_2(p)}}{n+1} \leq \binom{n}{\lfloor np \rfloor}\,. \tag{2.11}$$

For $0 < p \leq \frac{1}{2}$,

$$\frac{2^{nH_2(p)}}{n+1} \leq \binom{n}{\lceil np \rceil}\,. \tag{$\dagger$}$$

The proof uses the following lemma.

Lemma 2.1.37 *Let n be an integer and let $p \in (0,1)$ be such that np is an integer. Then*

$$\frac{2^{nH_2(p)}}{n+1} \leq \binom{n}{np} \leq 2^{nH_2(p)}\,.$$

Proof. The inequality

$$\binom{n}{np} p^{np}(1-p)^{n(1-p)} \leq 1$$

follows from the fact that the left-hand side is a probability, namely $P(\mathcal{B}(n,p) = np)$. Therefore

$$\binom{n}{np} \leq p^{-np}(1-p)^{-n(1-p)} = 2^{nH_2(p)}\,.$$

The integer value $k = np$ will be shown to maximize $\binom{n}{k}$ among all integers k such that $0 \leq k \leq n$. Therefore

$$\begin{aligned} 1 = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} &\leq (n+1)\binom{n}{np} p^{np}(1-p)^{n(1-p)} \\ &= (n+1)\binom{n}{np} 2^{-nH_2(p)}\,. \end{aligned}$$

To prove that $k = np$ maximizes $\binom{n}{k}$, compare two adjacent terms. We have

$$\begin{aligned} &\binom{n}{k} p^k (1-p)^{n-k} - \binom{n}{k+1} p^{k+1}(1-p)^{n-k-1} \\ &\qquad = \binom{n}{k} p^k (1-p)^{n-k} \left(1 - \frac{p(n-k)}{(1-p)(k+1)}\right)\,. \end{aligned}$$

This difference is non-negative if and only if

$$1 - \frac{p(n-k)}{(1-p)(k+1)} \geq 0$$

or, equivalently, $k \geq pn - (1-p)$. This shows that the function $k \to \binom{n}{k}$ increases as k varies from 0 to pn and decreases afterwards. □

We now proceed to the proof of Theorem 2.1.36:

Proof. Proof of (2.10):

$$\begin{aligned} \binom{n}{\lfloor np \rfloor} p^{pn}(1-p)^{(1-p)n} &\leq \binom{n}{\lfloor np \rfloor} p^{\lfloor np \rfloor}(1-p)^{n-\lfloor np \rfloor} \\ &\leq \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1\,. \end{aligned}$$

Inequality ($\star$) is proved in a similar way, or by an obvious symmetry argument. Inequality (2.11) follows from Lemma 2.1.37, since

$$\binom{n}{\lfloor np \rfloor} \geq \frac{2^{nH_2(\lfloor np \rfloor / n)}}{n+1} \geq \frac{2^{nH_2(p)}}{n+1}\,.$$

The proof of (†) is proved in a similar way, or by a symmetry argument. □

The Geometric Distribution

Definition 2.1.38 *A random variable X taking its values in $\mathbb{N}_+ := \{1, 2, \dots,\}$ and with the distribution*

$$P(T = k) = (1-p)^{k-1}p\,, \tag{2.12}$$

where $0 < p < 1$, is called a geometric random variable with parameter p.

This is denoted

$$X \sim \mathcal{G}\text{eo}(p)\,.$$

Of course, if $p = 1$, $P(T = 1) = 1$, and if $p = 0$, $P(T = \infty) = 1$. If $0 < p < 1$,

$$P(T < \infty) = \sum_{n=1}^{\infty}(1-p)^{k-1}p = \frac{1}{1-(1-p)} = \frac{p}{p} = 1\,,$$

and therefore $P(T = \infty) = 0$.

Example 2.1.39: First "heads" in the sequence. We are in the heads and tails framework. Define the random variable T to be the first time of occurrence of 1 in the sequence $X_1, X_2, \dots$, that is,

$$T = \inf\{n \geq 1; X_n = 1\},$$

with the convention that if $X_n = 0$ for all $n \geq 1$, then $T = \infty$. The event $\{T = k\}$ is exactly $\{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}$, and therefore,

$$P(T = k) = P(X_1 = 0) \cdots P(X_{k-1} = 0)P(X_k = 1),$$

that is, for $k \geq 1$,

$$P(T = k) = (1-p)^{k-1}p\,.$$

Theorem 2.1.40 *The mean of a geometric random variable X with parameter $p > 0$ is*

$$E[X] = \tfrac{1}{p}\,.$$

Proof.

$$E\left[X\right] = \sum_{k=1}^{\infty} k\left(1-p\right)^{k-1}p = \frac{1}{p^2} \times p = \frac{1}{p}\,.$$

□

Theorem 2.1.41 *A geometric random variable T with parameter $p \in (0,1)$ is memoryless in the sense that for any integers $k, k_0 \geq 1$, we have $P(T = k + k_0 \mid T > k_0) = P(T = k)$.*

Proof.

$$P\left(T>k_0\right)=\sum_{k=k_0+1}^{\infty}(1-p)^{k-1}p=(1-p)^{k_0}$$

and therefore

$$\begin{aligned}P\left(T=k_0+k|T>k_0\right)&=\frac{P\left(T=k_0+k,T>k_0\right)}{P\left(T>k_0\right)}=\frac{P\left(T=k_0+k\right)}{P\left(T>k_0\right)}\\&=\frac{p\left(1-p\right)^{k+k_0-1}}{(1-p)^{k_0}}=p\left(1-p\right)^{k}=P\left(T=k\right).\end{aligned}$$

□

Example 2.1.42: The coupon collector, take 1. In a certain brand of chocolate tablets one can find coupons, one in each tablet, randomly and independently chosen among n types. A prize may be claimed once the chocolate amateur has gathered a collection containing all the types of coupons. We seek to compute the average value of the number X of chocolate tablets bought when this happens for the first time.

For $0\le i\le n-1$, let X_i be the number of tablets it takes after ($>$) i different types of coupons have been collected to find a new type of coupon (in particular, $X_0=1$), so that

$$X=\sum_{i=0}^{n-1}X_i\,,$$

where each X_i ($1\le i\le n-1$) is a geometric random variable with parameter $p_i=1-\frac{i}{n}$. In particular,

$$E\left[X_i\right]=\frac{1}{p_i}=\frac{n}{n-i}\,,$$

(still true for $i=0$) and therefore

$$E\left[X\right]=\sum_{i=0}^{n-1}E\left[X_i\right]=n\sum_{i=0}^{n-1}\frac{1}{n-i}=n\sum_{i=1}^{n}\frac{1}{i}\,.$$

The sum $H(n):=\sum_{i=1}^{n}\frac{1}{i}$ (called the n-th harmonic number) satisfies the inequality

$$\log n\le H(n)\le\log n+1\,, \tag{2.13}$$

as can be seen by expressing $\log n$ as the integral $\int_1^n\frac{1}{x}\,dx$, partitioning the domain of integration with segments of unit length, and using the fact that the integrand is a decreasing function, which gives the inequalities

$$\sum_{i=2}^{n}\frac{1}{i}\le\int_1^n\frac{dx}{x}\le\sum_{i=1}^{n-1}\frac{1}{i}.$$

Therefore,

$$E\left[X\right]=(1+o(1))n\log n\,,$$

where $o(1)$ is a symbolic representation of a function of the positive integers that tend to 0 as $n\uparrow\infty$ (Landau's notation; see Section A.5).

The Hypergeometric Distribution

Recall Example 1.2.17. There is an urn containing N_1 black balls and N_2 red balls. You draw successively without replacement and at random n balls from the urn ($n \le N_1 + N_2$). The probability of having drawn k black balls ($0 \le k \le \inf(N_1, n)$) is:

$$p_k = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N_1+N_2}{n}}.$$

This probability distribution is called the hypergeometric distribution of parameters N_1 and N_2.

The Poisson Distribution

Definition 2.1.43 *A random variable X taking its values in $\mathbb{N}$ and such that for all $k \ge 0$,*

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

is called a Poisson random variable with parameter $\theta \ge 0$.

This is denoted by

$$X \sim \mathcal{P}\mathrm{oi}(\theta)\,.$$

If $\theta = 0$, $X \equiv 0$ (the general formula applies if one uses the convention $0! = 1$).

Example 2.1.44: The Poisson law of rare events, take 1. A veterinary surgeon in the Prussian cavalry once gathered data concerning the accidents due to horse kickbacks among soldiers. He deduced that the (random) number of accidents of the kind had a Poisson distribution. Here is an explanation.

Suppose that you play “heads and tails” for a large number n of (independent) tosses using a coin such that

$$P(X_i = 1) = \frac{\alpha}{n}.$$

In the Prussian army example, n is the (large) number of soldiers, and $X_i = 1$ if the i-th soldier has been hurt by a horse. Let S_n be the total number of heads (of wounded soldiers). We show that

$$\lim_{n\uparrow\infty} P(S_n = k) = e^{-\alpha}\frac{\alpha^k}{k!}\,, \tag{$\star$}$$

and this explains the findings of the veterinary surgeon. (The average number of casualties is α and the choice $P(X_i = 1) = \frac{\alpha}{n}$ guarantees this. Letting $n \uparrow \infty$ accounts for n being large but unknown.) Here is the proof of the mathematical statement.

The random variable S_n follows a binomial law with mean $n \times \frac{\alpha}{n} = \alpha$:

$$P(S_n = k) = \binom{n}{k}\left(\frac{\alpha}{n}\right)^k\left(1 - \frac{\alpha}{n}\right)^{n-k}.$$

In particular $P(S_n = 0) = \left(1 - \frac{\alpha}{n}\right)^n \to e^{-\alpha}$ as $n \uparrow \infty$. Also,

$$\frac{P(S_n = k+1)}{P(S_n = k)} = \frac{\frac{n-k}{k+1}\frac{\alpha}{n}}{1 - \frac{\alpha}{n}}$$

tends to $\frac{\alpha}{k+1}$ as $n \uparrow \infty$, from which $(\star)$ follows.

Theorem 2.1.45 *The mean of a Poisson random variable with parameter θ is given by*

$$E[X] = \theta\,,$$

and its variance is

$$\text{Var }(X) = \theta\,.$$

Proof.

$$E\left[X\right] = e^{-\theta}\sum_{k=1}^{\infty}\frac{\theta^k}{k!}k = e^{-\theta}\theta\sum_{j=0}^{\infty}\frac{\theta^j}{j!} = e^{-\theta}\theta e^{\theta} = \theta$$

and

$$\begin{aligned} E\left[X^2 - X\right] &= e^{-\theta}\sum_{k=0}^{\infty}\left(k^2 - k\right)\frac{\theta^k}{k!} = e^{-\theta}\sum_{k=2}^{\infty}k\left(k-1\right)\frac{\theta^k}{k!} \\ &= e^{-\theta}\theta^2\sum_{k=2}^{\infty}\frac{\theta^{k-2}}{(k-2)!} = e^{-\theta}\theta^2\sum_{j=0}^{\infty}\frac{\theta^j}{j!} = e^{-\theta}\theta^2 e^{\theta} = \theta^2\,. \end{aligned}$$

Therefore

$$\begin{aligned} \text{Var }\ (X) &= E\left[X^2\right] - E\left[X\right]^2 \\ &= E\left[X^2 - X\right] + E\left[X\right] - E\left[X\right]^2 = \theta^2 + \theta - \theta^2 = \theta. \end{aligned}$$

□

Theorem 2.1.46 *Let X_1 and X_2 be two independent Poisson random variables with means $\theta_1 > 0$ and $\theta_2 > 0$, respectively. Then $X = X_1 + X_2$ is a Poisson random variable with mean $\theta = \theta_1 + \theta_2$.*

Proof. For $k \geq 0$,

$$\begin{aligned} P(X = k) = P(X_1 + X_2 = k) &= P\left(\sum_{i=0}^{k}\{X_1 = i, X_2 = k - i\}\right) \\ &= \sum_{i=0}^{k}P(X_1 = i, X_2 = k-i) = \sum_{i=0}^{k}P(X_1 = i)P(X_2 = k-i) \\ &= \sum_{i=0}^{k}e^{-\theta_1}\frac{\theta_1^i}{i!}e^{-\theta_2}\frac{\theta_2^{k-i}}{(k-i)!} = e^{-(\theta_1+\theta_2)}\frac{(\theta_1+\theta_2)^k}{k!}, \end{aligned}$$

where we used the binomial formula. □

The Multinomial Distribution

Consider the random vector $X = (X_1, \dots, X_N)$ where all the random variables X_i take their values in the *same* (this restriction is not essential, but it simplifies the notation) denumerable space E. Let $p : E^N \to \mathbb{R}_+$ be a function such that

$$\sum_{x \in E^N} p(x) = 1 .$$

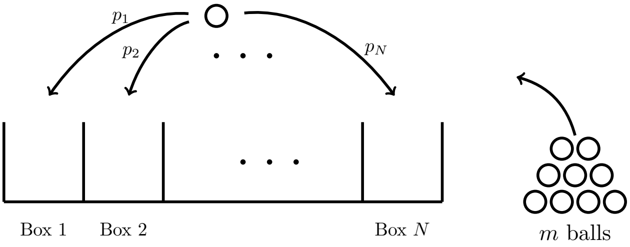
Definition 2.1.47 *The discrete random vector X above is said to admit the probability distribution p if for all sets $C \subseteq E^N$,*

$$P(X \in C) = \sum_{x \in C} p(x) .$$

In fact, there is nothing new here since X is a discrete random variable taking its values in the denumerable set $\mathcal{X} := E^N$.

Consider m balls to be placed in N boxes $B_1, \dots, B_N$ independently of one another, with the probability p_i for a given ball to be assigned to box B_i. Of course,

$$\textstyle\sum_{i=1}^N p_i = 1 .$$



After placing all the balls in the boxes, there are X_i balls in box B_i, where

$$\textstyle\sum_{i=1}^N X_i = m .$$

The random vector $X = (X_1, \dots, X_N)$ is a multinomial vector of size (N, m) and parameters $p_1, \dots, p_N$, that is, its probability distribution is

$$P(X_1 = m_1, \dots, X_N = m_N) = \tfrac{k!}{\prod_{i=1}^N (m_i)!} \textstyle\prod_{i=1}^N p_i^{m_i} ,$$

where $m_1 + \cdots + m_N = m$.

Proof. Observe that (α): there are $m!/\prod_{i=1}^N (m_i)!$ distinct ways of placing m balls in N boxes in such a manner that m_1 balls are in box B_1, m_2 are in B_2, etc., and (β): each of these distinct ways occurs with the same probabilty $\prod_{i=1}^N p_i^{m_i}$. □

The Uniform Distribution on $[0,1]$

This subsection introduces non-discrete random variables. In fact, it gives just what is strictly necessary in this book, in particular, the notion of independent random numbers.

Definition 2.1.48 *A function $X:\Omega\to\mathbb{R}$ such that for all $x\in\mathbb{R}$*

$$\{\omega; X(\omega)\le x\}\in\mathcal{F}$$

is called a real random variable.

Its cumulative distribution function is the function $F(x):=P(X\le x)$. If

$$F(x)=\int_{-\infty}^{x} f(y)\,dy\,,$$

for all $x\in\mathbb{R}$ for some non-negative function f such that $\int_{-\infty}^{+\infty} f(y)\,dy=1$, the latter is called the probability density function, or PDF, of X.

The following example is all we need in this book.

EXAMPLE 2.1.49: THE UNIFORM DISTRIBUTION. Let $[a,b]\in\mathbb{R}$. A real random variable X with the PDF

$$f(x)=\frac{1}{b-a}1_{[a,b]}(x)$$

is called a uniform random variable on $[a,b]$. This is denoted by

$$X\sim\mathcal{U}([a,b])\,.$$

Uniform random variables are used in simulation, more precisely, to generate a discrete random variable Z with a prescribed distribution $P(Z=a_i)=p_i\quad(0\le i\le K)$. The basic principle of the sampling algorithm is the following

Draw $U\sim\mathcal{U}([0,1])$.

Set $Z=a_\ell$ if $U\in I_\ell:=(p_0+p_1+\ldots+p_{\ell-1},p_0+p_1+\ldots+p_\ell]$.

Indeed, since the interval I_ℓ has length p_l, $P(Z=a_\ell)=P(U\in I_\ell)=p_\ell$.

This method is called the method of the inverse.

Definition 2.1.50 *A real random vector of dimension d is a mapping $X=(X_1,\ldots,X_d):\Omega\to\mathbb{R}$ such that each coordinate X_i is a real random variable.*

A non-negative function $f:\mathbb{R}^d\to\mathbb{R}$ such that $\int_{\mathbb{R}^d} f(x)\,dx=1$ and

$$P(X_1\le x_1,\ldots,X_d\le x_d)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_d} f(x_1,\ldots,x_d)\,dx_1\cdots dx_d$$

is called the probability distribution function of the random vector X.

Definition 2.1.51 *The real random variables $X_1, \dots, X_d$ admitting the respective* PDF*'s $f_1, \dots, f_d$ are said to be independent if the* PDF *of the random vector $X = (X_1, \dots, X_d)$ is of the form*

$$f(x_1, \dots, x_d) = f_1(x_1) \times \cdots \times f(x_d)$$

where the f_i's are non-negative functions such that $\int_{-\infty}^{+\infty} f_i(y)\, dy = 1$.

The f_i's are then the PDF's of the X_i's. For instance with $i = 1$,

$$\begin{aligned} P(X_1 \le x_1) &= P(X_1 \le x_1, X_2 < \infty \dots, X_d < \infty) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_1(x_1) f_2(x_2) \cdots f_d(x_d)\, dx_1 \cdots dx_d \\ &= \int_{-\infty}^{x_1} f_1(x_1)\, dx_1 \int_{-\infty}^{+\infty} f_2(x_2)\, dx_2 \cdots \int_{-\infty}^{+\infty} f_d(x_d) dx_d = \int_{-\infty}^{x_1} f_1(x_1)\, dx_1 \,. \end{aligned}$$

Definition 2.1.52 *The real random variables $X_1, X_2 \dots$ admitting the respective* PDF*'s $f_1, f_2, \dots$ are said to be independent if for all integers $k \ge 2$, the random variables $X_1, \dots, X_d$ are independent.*

EXAMPLE 2.1.53: SEQUENCE OF INDEPENDENT RANDOM NUMBERS. The sequence $\{U_n\}_{n \ge 1}$ is called a sequence of independent random numbers if for all $k \ge 1$, $U_1, \dots, U_k$ are independent random variables uniformly distributed on the interval $[0, 1]$.

The Gilbert–Erdös–Rényi Random Graphs

A graph is a discrete object and therefore random graphs are, from the purely formal point of view, discrete random variables. The random graphs considered in this book are in fact described by a finite collection of IID $\{0, 1\}$-valued random variables. They will be studied in more detail in Chapter 10. The basic definitions of graph theory below will be complemented as the need arises.

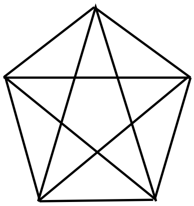
A (finite) graph $(V, \mathcal{E})$ consists of a finite collection V of vertices v and of a collection $\mathcal{E}$ of unordered pairs of distinct vertices, $\langle u, v \rangle$, called the edges. If $\langle u, v \rangle \in \mathcal{E}$, then u and v are called neighbours, and this is also denoted by $u \sim v$. The degree of vertex $v \in V$ is the number of edges stemming from it.

In a few occasions, some redundancy in the notation will be useful: V and $\mathcal{E}$ will be denoted by $V(G)$ and $\mathcal{E}(G)$.

A subgraph (or induced subgraph) of a graph $G = (V, \mathcal{E})$ is any graph $G' = (V', \mathcal{E}')$ with $V' \subseteq V$ and $\mathcal{E}' = \{\langle u, v \rangle \in \mathcal{E};\, u, v \in V'\}$. Such graph is also called the restriction of G to V' and is denoted by $G|_{V'}$.

A complete graph is one having all the possible $\binom{n}{2}$ edges. It will be denoted by K_n and its edge set by $\mathbf{E_n}$. Note that a subgraph of a complete graph is also complete.

A complete subgraph is called a clique of the graph. Note that a singleton of V is a clique.



The complete pythagorean graph

A graph is connected if for all pairs of distinct vertices v, w, there is a sequence $v_0 = v, v_1, \dots, v_n = w$ (called a path from v to w) and such that $v_0 \sim v_1 \sim \dots \sim v_n$.

A cycle of a graph is a sequence of distinct vertices $v_1, v_2, \dots, v_n$ such that $v_1 \sim v_2 \sim \dots \sim v_n \sim v_1$. A tree is a connected graph without cycles.

Let $G_1 = (V, \mathcal{E}_1)$ and $G_2 = (V, \mathcal{E}_2)$ be two graphs with the same set of vertices. The graph $G = G_1 \cup G_2$ is by definition the graph on the set of vertices V such that $e \in \mathcal{E}$ if and only if $e \in \mathcal{E}_1 \cup \mathcal{E}_2$. This graph is called the union of G_1 and G_2. One defines similarly the intersection of G_1 and G_2, $G = G_1 \cap G_2$, to be the graph on the set of vertices V such that $e \in \mathcal{E}$ if and only if $e \in \mathcal{E}_1 \cap \mathcal{E}_2$. One writes $G_2 \subseteq G_1$ if and only if $\mathcal{E}(G_1) \subseteq \mathcal{E}(G_2)$.

Some graph properties may be difficult to verify on a given graph. However, there exist results showing that they are satisfied (or not) for “large” and “typical” graphs. The question of course is: what is a typical graph? One possible choice is the Gilbert random graph (Definition 2.1.54 below).

Definition 2.1.54 *(Gilbert, 1959) Let n be a fixed positive integer and let $V = \{1, 2, \dots, n\}$ be a finite set of vertices. To each unordered pair of distinct vertices $\langle u, v\rangle$, associate a random variable $X_{\langle u,v\rangle}$ taking its values in $\{0, 1\}$ and suppose that all such variables are* IID *with probability $p \in (0, 1)$ for the value 1. This defines a random graph denoted by $\mathcal{G}(n, p)$, a random element taking its values in the (finite) set of all graphs with vertices $\{1, 2, \dots, n\}$ and admitting for edge the unordered pair of vertices $\langle u, v\rangle$ if and only if $X_{\langle u,v\rangle} = 1$.*

Note that $\mathcal{G}(n, p)$ is indeed a discrete random variable (taking its values in the finite set consisting of the collection of graphs with vertex set $V = \{1, 2, \dots, n\}$). Similarly, the set $\mathcal{E}_{n,p}$ of edges of $\mathcal{G}(n, p)$ is also a discrete random variable. If we call any unordered pair of vertices $\langle u, v\rangle$ a potential edge (there are $\binom{n}{2}$ such edges forming the set $\mathbf{E_n}$), $\mathcal{G}(n, p)$ is constructed by accepting a potential edge as one of its edges with probability p independently of all other potential edges. The probability of occurence of a graph G with exactly m edges is then

$$P(\mathcal{G}(n, p) = G) = P(|\mathcal{E}_{n,p}| = m) = p^m (1 - p)^{\binom{n}{2} - m}\,.$$

Note that the degree of a given vertex, that is the number of edges stemming from it, is a binomial random variable $\mathcal{B}(n-1,p)$. In particular, the average degree is $d=(n-1)p$.

Another type of random graph is the Erdös–Rényi random graph (Definition 2.1.55 below). It is closely related to the Gilbert graph as we shall see below, in Theorem 2.1.56.

Definition 2.1.55 *(Erdös and Rényi, 1959) Consider the collection* $\mathbf{G_m}$ *of graphs* $G=(V,\mathcal{E})$ *where* $V=\{1,2,\dots,n\}$ *with exactly* m *edges* ($|\mathcal{E}|=m$)*. There are* $\binom{\binom{n}{2}}{m}$ *such graphs. The Erdös–Rényi random graph* $\mathcal{G}_{n,m}$ *is a random graph uniformly distributed on* $\mathbf{G_m}$.

(The notation is chosen for a quick differentiation between Gilbert graphs $\mathcal{G}_{n,m}$ and Erdös–Rényi graphs $\mathcal{G}(n,p)$.)

Denoting by $\mathcal{E}_{n,m}$ the (random) collection of edges of $\mathcal{G}_{n,m}$, the probability of obtaining a given graph $G\in\mathbf{G_m}$ is

$$P(G)=\binom{\binom{n}{2}}{m}^{-1}.$$

The random graph $\mathcal{G}_{n,m}$ can be constructed by including m edges successively at random. More precisely, denoting by G_k $(0\le k\le m)$ the successive graphs, and by $\mathcal{E}_k$ the collection of edges of G_k, $G_0=(V,\varnothing)$ and for $1\le k\le m$, $\mathcal{E}_k=\mathcal{E}_{k-1}\cup e_k$, where

$$P(e_k=e\mid G_0,\dots,G_{k-1})=|\mathbf{E_n}\backslash\mathcal{E}_{k-1}|^{-1}$$

for all edges $e\in\mathbf{E_n}\backslash\mathcal{E}_{k-1}$.

Theorem 2.1.56 *The conditional distribution of* $\mathcal{G}(n,p)$ *given that the numner of edges is* $m\le\binom{n}{2}$ *is uniform on the set* $\mathbf{G_m}$ *of graphs* $G=(V,\mathcal{E})$ *where* $V=\{1,2,\dots,n\}$ *with exactly* m *edges.*

Proof. Let G be a graph with vertex set V have exactly m edges. Observing that $\{\mathcal{G}(n,p)=G\}\subseteq\{|\mathcal{E}_{n,p}|=m\}$, we have that

$$\begin{aligned}P(\mathcal{G}(n,p)=G\mid|\mathcal{E}_{n,p}|=m)&=\frac{P(\mathcal{G}(n,p)=G,|\mathcal{E}_{n,p}|=m)}{P(|\mathcal{E}_{n,p}|=m)}\\&=\frac{P(\mathcal{G}(n,p)=G)}{P(|\mathcal{E}_{n,p}|=m)}\\&=\frac{p^m(1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m}p^m(1-p)^{\binom{n}{2}-m}}=\binom{\binom{n}{2}}{m}^{-1}.\end{aligned}$$

□

Remark 2.1.57 In the sequel, we will follow the tradition of refering to Gilbert graphs as Erdös–Rényi graphs.

2.2 Generating functions

2.2.1 Definition and Properties

The computation of probabilities in discrete probability models often require an enumeration of all the possible outcomes realizing this particular event. Generating functions are very useful for this task, and more generally, for obtaining the probability distributions of integer-valued random variables. We first define the expectation of a complex-valued function of a random variable.

Let X be a discrete random variable with values in $\mathbb{N}$, and let $\varphi : \mathbb{N} \to \mathbb{C}$ be a complex function with real and imaginary parts φ_R and φ_I respectively. The expectation $E[\varphi(X)]$ is naturally defined by

$$E[\varphi(X)] := E[\varphi_R(X)] + iE[\varphi_I(X)]\,,$$

provided the expectations on the right-hand side are well-defined and finite. This is the case if $E\left[|\varphi(X)|\right] < \infty$.

Definition 2.2.1 *Let X be an integer-valued random variable. Its generating function (GF) is the function $g : \mathcal{D} \to \mathbb{C}$ defined by*

$$g(z) := E[z^X] = \sum_{k=0}^{\infty} P(X = k)z^k\,, \tag{2.14}$$

and where $\mathcal{D} := \overline{D}(0; R) := \{z \in \mathbb{C}; |z| \leq R\}$ is the closed disk of absolute convergence of the above series.

Since $\sum_{n=0}^{\infty} P(X = n) = 1 < \infty$, $R \geq 1$. In the next two examples, $R = \infty$.

EXAMPLE 2.2.2: GF OF THE BINOMIAL VARIABLE. For the binomial random variable of size n and parameter p,

$$g(z) = \textstyle\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = \sum_{k=0}^{n} \binom{n}{k} (zp)^k (1-p)^{n-k}\,,$$

and therefore

$$g(z) = (1 - p + pz)^n\,.$$

EXAMPLE 2.2.3: GF OF THE POISSON VARIABLE. For the Poisson random variable of mean θ,

$$g(z) = e^{-\theta} \textstyle\sum_{k=0}^{\infty} \frac{(\theta)^k}{k!} z^k = e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta z)^k}{k!},$$

and therefore

$$g(z) = e^{\theta(z-1)}\,.$$

Next is an example where the radius of convergence is finite.

EXAMPLE 2.2.4: GF OF THE GEOMETRIC VARIABLE. For the geometric random variable of (2.12),

$$g(z) = \textstyle\sum_{k=0}^{\infty} p(1-p)^{k-1}\, z^k \,.$$

The radius of convergence of this generating function power series is $\frac{1}{1-p}$ and its sum is

$$g(z) = \textstyle\sum_{k=0}^{\infty} pz((1-p)z)^{k-1} = \frac{pz}{1-qz} \,.$$

Theorem 2.2.5 *The generating function characterizes the distribution of a random variable.*

This means the following. Suppose that, without knowing the distribution of X, you have been able to compute its generating function g, and that, moreover, you are able to give its power series expansion in a neighborhood of the origin[1], say,

$$g(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Since g is the generating function of X,

$$g(z) = \sum_{n=0}^{\infty} P(X=n) z^n$$

and since the power series expansion around the origin is unique, $P(X=n) = a_n$ for all $n \geq 0$. Similarly, if two integer-valued random variables X and Y have the same generating function, they have the same distribution. Indeed, the identity in a neighborhood of the origin of two power series implies the identity of their coefficients.

Theorem 2.2.6 *Let X and Y be two independent integer-valued random variables with respective generating functions g_X and g_Y. Then the sum $X+Y$ has the* GF

$$g_{X+Y}(z) = g_X(z) \times g_Y(z).$$

Proof. Use the product formula for expectations:

$$g_{X+Y}(z) = E\left[z^{X+Y}\right] = E\left[z^X z^Y\right] = E\left[z^X\right] E\left[z^Y\right] \,.$$

□

EXAMPLE 2.2.7: SUM OF INDEPENDANT POISSON VARIABLES. Let X and Y be two *independent* Poisson random variables with means α and β respectively. The

[1] This is a common situation; see Theorem 2.2.10 for instance.

sum $X+Y$ is a Poisson random variable with mean $\alpha+\beta$. Indeed, by Theorem 2.2.6,

$$g_{X+Y}(z) = g_X(z) \times g_Y(z) = e^{\alpha(z-1)}\, e^{\beta(z-1)} = e^{(\alpha+\beta)(z-1)},$$

and the assertion follows directly from Theorem 2.2.5 since g_{X+Y} is the GF of a Poisson random variable with mean $\alpha+\beta$.

The next result gives concerns the shape of the generating function restricted to the interval $[0,1]$.

Theorem 2.2.8 *(α) Let $g : [0,1] \to \mathbb{R}$ be defined by $g(x) = E[x^X]$, where X is a non-negative integer-valued random variable. Then g is nondecreasing and convex. Moreover, if $P(X=0) < 1$, it is strictly increasing, and if $P(X \le 1) < 1$, it is strictly convex.*

(β) Suppose $P(X \le 1) < 1$. If $E[X] \le 1$, the equation $x = g(x)$ has a unique solution $x \in [0,1]$, namely $x = 1$. If $E[X] > 1$, it has two solutions in $[0,1]$, $x = 1$ and $x = x_0 \in (0,1)$.

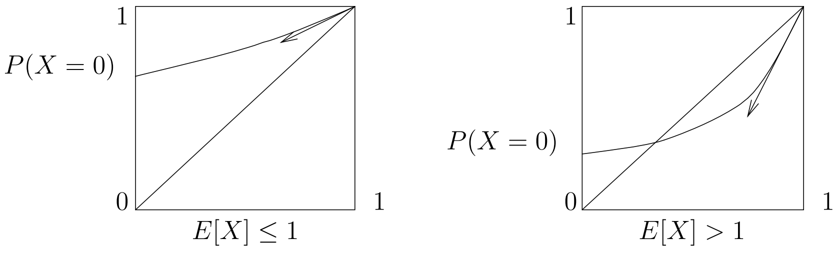
Proof. Just observe that for $x \in [0,1]$,

$$g'(x) = \sum_{n=1}^{\infty} nP(X=n)x^{n-1} \ge 0,$$

and therefore g is nondecreasing, and

$$g''(x) = \sum_{n=2}^{\infty} n(n-1)P(X-n)x^{n-2} \ge 0,$$

and therefore g is convex. For $g'(x)$ to be null for some $x \in (0,1)$, it is necessary to have $P(X=n) = 0$ for all $n \ge 1$, and therefore $P(X=0) = 1$. For $g''(x)$ to be null for some $x \in (0,1)$, one must have $P(X=n) = 0$ for all $n \ge 2$, and therefore $P(X=0) + P(X=1) = 1$.



$E[X] \le 1$

$E[X] > 1$

Two aspects of the generating function

The graph of $g : [0,1] \to \mathbb{R}$ has, in the strictly increasing strictly convex case $P(X=0) + P(X=1) < 1$, the general shape shown in the figure, where we distinguish two cases: $E[X] = g'(1) \le 1$, and $E[X] = g'(1) > 1$. The rest of the proof is then easy. □

Moments from the Generating Function

Generating functions are powerful computational tools. First of all, they can be used to obtain moments of a discrete random variable.

Theorem 2.2.9 *We have*

$$g'(1) = E[X] \tag{2.15}$$

and

$$g''(1) = E[X(X-1)]. \tag{2.16}$$

Proof. Inside the open disk $D(0; R)$ centered at the origin and of radius R, the power series defining the generating function g is continuous, and differentiable at any order term by term. In particular, differentiating twice both sides of (2.14) inside the open disk $D(0; R)$ gives

$$g'(z) = \sum_{n=1}^{\infty} nP(X = n)z^{n-1}, \tag{2.17}$$

and

$$g''(z) = \sum_{n=2}^{\infty} n(n-1)P(X = n)z^{n-2}. \tag{2.18}$$

When the radius of convergence R is *strictly larger* than 1, we obtain the announced results by letting $z = 1$ in the previous identities.

If $R = 1$, the same is basically true but the mathematical argument is more subtle. The difficulty is not with the right-hand side of (2.17), which is always well-defined at $z = 1$, being equal to $\sum_{n=1}^{\infty} nP(X = n)$, a non-negative and possibly infinite quantity. The difficulty is that g may be not differentiable at $z = 1$, a boundary point of the disk (here of radius 1) on which it is defined. However, by *Abel's theorem* (Theorem A.1.3), the limit as the *real* variable x increases to 1 of $\sum_{n=1}^{\infty} nP(X = n)x^{n-1}$ is $\sum_{n=1}^{\infty} nP(X = n)$. Therefore g', as a function on the real interval $[0, 1)$, can be extended to $[0, 1]$ by (2.15), and this extension preserves continuity. With this *definition* of $g'(1)$, Formula (2.15) holds true. Similarly, when $R = 1$, the function g'' defined on $[0, 1)$ by (2.18) is extended to a continuous function on $[0, 1]$ by *defining* $g''(1)$ by (2.16). □

2.2.2 Random Sums

How to compute the distribution of random sums? Here again, generating functions help.

Theorem 2.2.10 *Let $\{Y_n\}_{n\geq 1}$ be an* IID *sequence of integer-valued random variables with the common generating function g_Y. Let T be another random variable, integer-valued, independent of the sequence $\{Y_n\}_{n\geq 1}$, and let g_T be its generating function. The generating function of*

$$X = \sum_{n=1}^{T} Y_n \,,$$

where by convention $\sum_{n=1}^{0} = 0$, *is*

$$g_X(z) = g_T\left(g_Y(z)\right) \,. \tag{2.19}$$

Proof.

$$\begin{aligned}
\sum_{n\geq 0} z^n P(X=n) &= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(X=n, T=k\right)\right)\\
&= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(\sum_{j=1}^{k} Y_j = n, T=k\right)\right)\\
&= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(\sum_{j=1}^{k} Y_j = n, T=k\right)\right)\\
&= \sum_{n\geq 0} z^n \left(\sum_{k\geq 0} P\left(\sum_{j=1}^{k} Y_j = n\right) P\left(T=k\right)\right)\\
&= \sum_{k\geq 0} P\left(T=k\right) \left(\sum_{n\geq 0} z^n P\left(\sum_{j=1}^{k} Y_j = n\right)\right) \,.
\end{aligned}$$

But

$$\left(\sum_{n\geq 0} z^n P\left(\sum_{j=1}^{k} Y_j = n\right)\right) = g_{\sum_{j=1}^{k} Y_j}(z) = \left(g_Y(z)\right)^k \,.$$

Therefore,

$$\sum_{n\geq 0} z^n P(X=n) = \sum_{k\geq 0} P\left(T=k\right)\left(g_Y(z)\right)^k = g_T\left(g_Y(z)\right) \,.$$

□

By taking derivatives in (2.19),

$$E\left[X\right] = g'_X(1) = g'_Y(1) g'_T\left(g_Y(1)\right) = E[Y_1]E[T].$$

This is Wald's formula. Exercise 2.4.16 gives more general conditions for its validity.

2.2.3 Counting with Generating Functions

The following example is typical of the use of generating functions in combinatorics (the art of counting).

EXAMPLE 2.2.11: LOTTERY. Let X_1, X_2, X_3, X_4, X_5, and X_6 be independent random variables uniformly distributed over $\{0, 1, \ldots, 9\}$. We shall compute the generating function of $Y = 27 + X_1 + X_2 + X_3 - X_4 - X_5 - X_6$ and use the result to obtain the probability that in a 6-digit lottery the sum of the first three digits equals the sum of the last three digits. We have

$$E[z^{X_i}] = \frac{1}{10}(1 + z + \cdots + z^9) = \frac{1}{10}\frac{1 - z^{10}}{1 - z},$$

and therefore

$$E[z^{-X_i}] = \frac{1}{10}\frac{1}{z^9}\frac{1 - z^{10}}{1 - z},$$

and

$$\begin{aligned} E[z^Y] &= E[\left[z^{27+\sum_{i=1}^{3} X_i - \sum_{i=4}^{6} X_i}\right] \\ &= E\left[z^{27}\prod_{i=1}^{3} z^{X_i}\prod_{i=4}^{6} z^{-X_i}\right] = z^{27}\prod_{i=1}^{3} E[z^{X_i}]\prod_{i=4}^{6} E[z^{-X_i}]\,. \end{aligned}$$

Therefore,

$$g_Y(z) = \frac{1}{10^6}\frac{(1 - z^{10})^6}{(1 - z)^6}.$$

But $P(X_1 + X_2 + X_3 = X_4 + X_5 + X_6) = P(Y = 27)$ is the factor of z^{27} in the power series expansion of $g_Y(z)$. Since

$$(1 - z^{10})^6 = 1 - \binom{6}{1} z^{10} + \binom{6}{2} z^{20} + \cdots$$

and

$$(1 - z)^{-6} = 1 + \binom{6}{5} z + \binom{7}{5} z^2 + \binom{8}{5} z^3 + \cdots$$

(recall the negative binomial formula:

$$(1 - z)^{-p} = 1 + \binom{p}{p-1} z + \binom{p+1}{p-1} z^2 + \binom{p+2}{p-1} z^3 + \cdots),$$

we find that

$$P(Y = 27) = \frac{1}{10^6}\left(\binom{32}{5} - \binom{6}{1}\binom{22}{5} + \binom{6}{2}\binom{12}{5}\right).$$

2.3 Conditional Expectation

2.3.1 Conditioning with Respect to an Event

Chapter 1 introduced the notion of conditional probability and the Bayes calculus associated with it. We now introduce the notion of conditional expectation and the set of rules accompanying it.

Let Z be a discrete random variable with values in E, and let $f : E \to \mathbb{R}$ be a non-negative function. Let A be some event of positive probability. The conditional expectation of $f(Z)$ given A, denoted by $E\left[f(Z)\,|\,A\right]$, is by definition the expectation when the distribution of Z is replaced by its conditional distribution given A, $P(Z = z\,|\,A)$. Therefore

$$E\left[f(Z)\,|\,A\right] := \sum_z f(z)P(Z = z\,|\,A)\,.$$

Let $\{A_i\}_{i\in\mathbb{N}}$ be a partition of the sample space. Then

$$E\left[f(Z)\right] = \sum_{i\in\mathbb{N}} E\left[f(Z)\,|\,A_i\right] P(A_i)\,.$$

Proof. This is a direct consequence of the Bayes formula of total causes:

$$\begin{aligned} E\left[f(Z)\right] &= \sum_z f(z)P(Z = z) = \sum_z \left(\sum_i f(z)P(Z = z\,|\,A_i)P(A_i)\right) \\ &= \sum_i \left(\sum_z f(z)P(Z = z\,|\,A_i)\right) P(A_i) = \sum_i E\left[f(Z)\,|\,A_i\right] P(A_i)\,. \end{aligned}$$

□

The following elementary result will often be used, and therefore, we shall promote it to the rank of theorem:

Theorem 2.3.1 *Let Z be a random variable with values in E, and let $f : E \mapsto \mathbb{R}$ be a non-negative function. Let A be some event of positive probability. Then*

$$E\left[f(Z)1_A\right] = E\left[f(Z)\,|\,A\right] P(A)\,.$$

Proof.

$$E\left[f(Z)\,|\,A\right] P(A) = \left(\sum_{z\in E} f(z)P(Z = z\,|\,A)\right) P(A) = \sum_{z\in E} f(z)P(Z = z\,,\,A)\,.$$

Now, the random variable $f(Z)1_A$ takes a non-null value if and only if this value is of the form $f(z) > 0$, and this happens with probability $P(Z = z\,,\,A)$. Therefore

$$E\left[f(Z)1_A\right] = \sum_{z\,;f(z)>0} f(z)P(Z = z\,,\,A) = \sum_{z\in E} f(z)P(Z = z\,,\,A)\,.$$

□

EXAMPLE 2.3.2: POISSON BOUNDING OF MULTINOMIAL EVENTS. (Mitzenmacher and Upfal, 2005.) The computation of expectations concerning multinomial vectors often turns out to be difficult, whereas it might be considerably simpler in the Poisson case. The result of this subsection gives, under certain conditions, a bound for the expectation of interest in terms of the expectation computed for the Poisson case. Before the precise statement of this result, some preliminary remarks are in order.

Balls are placed in N bins in the following manner. The number of balls in any given bin is a Poisson variable of mean $\frac{m}{N}$, and is independent of the numbers in the other bins. In particular, the total number of balls $Y_1 + \cdots + Y_N$ is, as the sum of independent Poisson random variables, a Poisson random variable whose mean is the sum of the means of the coordinates, that is m.

Let $f \geq 0$ be a function of N integer-valued arguments, and let $(X_1, \ldots, X_N)$ be a multinomial random vector of size (m, N) and with parameters $p_i = \frac{1}{N}$ (obtained by placing m balls independently and at random in N bins). Then, with the Y_i's as above,

$$E\left[f(X_1, \ldots, X_N)\right] \leq e\sqrt{m} E\left[f(Y_1, \ldots, Y_N)\right] . \tag{2.20}$$

In particular, with f the indicator of some subset E of $\mathbb{N}^N$, the probability that $(X_1, \ldots, X_N) \in E$ is less than $e\sqrt{m}$ times the probability that $(Y_1, \ldots, Y_N) \in E$. This can be rephrased in imprecise but suggestive terms as follows: An event that has probability P in the Poisson case happens with probability at most $e\sqrt{m}P$ in the multinomial case.

Proof. For a given arbitrary integer k, the conditional probability that there are k_1 balls in bin 1, k_2 balls in bin 2, ..., given that the total number of balls is $k_1 + \cdots + k_N = k$ is

$$\begin{aligned} & P\left(Y_1 = k_1, \ldots, Y_N = k_N \,|\, Y_1 + \cdots + Y_N = k\right) \\ & \qquad = \frac{P\left(Y_1 = k_1, \ldots, Y_N = k_N, Y_1 + \cdots + Y_N = k\right)}{P\left(Y_1 + \cdots + Y_N = k\right)} \\ & \qquad = \frac{P\left(Y_1 = k_1, \ldots, Y_N = k_N\right)}{P\left(Y_1 + \cdots + Y_N = k\right)} . \end{aligned}$$

By independence of the Y_i's and since they are Poisson variables with mean $\frac{m}{N}$,

$$P\left(Y_1 = k_1, \ldots, Y_N = k_N\right) = \prod_{i=1}^{N} \left(e^{-\frac{m}{N}} \frac{\left(\frac{m}{N}\right)^{k_i}}{k_i!} \right) .$$

Also, $P\left(Y_1 + \cdots + Y_N = k\right) = e^{-m}\frac{m^k}{k!}$. Therefore

$$P\left(Y_1 = k_1, \ldots, Y_N = k_N \,|\, Y_1 + \cdots + Y_N = k\right) = \frac{k!}{k_1! \cdots k_N!} \left(\frac{1}{N}\right)^N .$$

But this is equal to $P(Z_1 = k_1, \ldots, Z_N = k_N)$, where Z_i is the number of balls in bin i when $k = k_1 + \cdots + k_N$ balls are placed independently and at random in the N bins. Note that the above equality is independent of m.

Now:

$$
\begin{aligned}
E\left[f(Y_1,\dots,Y_N)\right] &= \sum_{k=0}^{\infty} E\left[f(Y_1,\dots,Y_N)\mid \sum_{i=1}^{N} Y_i = k\right] P\left(\sum_{i=1}^{N} Y_i = k\right) \\
&\geq E\left[f(Y_1,\dots,Y_N)\mid \sum_{i=1}^{N} Y_i = m\right] P\left(\sum_{i=1}^{N} Y_i = m\right) \\
&= E\left[f(X_1,\dots,X_N)\right] P\left(\sum_{i=1}^{N} Y_i = m\right) \\
&= E\left[f(X_1,\dots,X_N)\right] \frac{m^m e^{-m}}{m!}.
\end{aligned}
$$

The announced result will follow from the bound

$$
m! \leq e\sqrt{m}\left(\frac{m}{e}\right)^m . \qquad (\star)
$$

For this, use the fact that, by concavity of the function $x \to \log x$,

$$
\int_{i-1}^{i} \log x\, dx \geq \frac{\log(i-1)+\log i}{2},
$$

and therefore

$$
\int_{1}^{m} \log x\, dx \geq \sum_{i=1}^{m} \log i - \frac{\log m}{2} = \log(m!) - \frac{\log m}{2} .
$$

Integration by parts gives $m\log m - m + 1 = \int_1^m \log x\, dx$. Therefore $m\log m - m + 1 \geq \log(m!) - \frac{\log m}{2}$, from which the announced inequality follows by taking exponentials. □

There exists a stronger version of (2.20):

$$
E\left[f(X_1,\dots,X_N)\right] \leq 4E\left[f(Y_1,\dots,Y_N)\right] ,
$$

but this time it is required in addition that $E\left[f(X_1,\dots,X_N)\right]$ should be a quantity increasing with the number m of balls.

Proof.

$$
\begin{aligned}
E\left[f(Y)\right] &= \sum_{k=0}^{\infty} E\left[f(Y)\mid \sum Y_i = k\right] P\left(\sum Y_i = k\right) \\
&\geq \sum_{k=m}^{\infty} E\left[f(Y)\mid \sum Y_i = k\right] P\left(\sum Y_i = k\right) \\
&\geq E\left[f(Y)\mid \sum Y_i = m\right] P\left(\sum Y_i = k\right) \\
&\geq E\left[f(X)\right]\; P\left(\sum Y_i = k\right) \geq E\left[f(X)\right] \times \frac{1}{4} ,
\end{aligned}
$$

since for any Poisson variable Z with a mean θ that is a positive integer, $P(Z \geq \theta) \geq \frac{1}{4}$. □

2.3.2 Conditioning with Respect to a Random Variable

Let X and Y be two discrete random variables taking their values in the denumerable sets F and G respectively. Let the function $g : F \times G \to \mathbb{R}$ be either non-negative, or such that $E[|g(X,Y)|] < \infty$. For each $y \in G$ such that $P(Y = y) > 0$, define

$$\psi(y) := \sum_{x \in F} g(x,y) P(X = x \,|\, Y = y)\,, \tag{2.21}$$

and let $\psi(y) := 0$ otherwise. The sum in (2.21) is well-defined (possibly infinite however) when g is non-negative. Note that in the non-negative case, we have that

$$\begin{aligned}\sum_{y \in G} \psi(y) P(Y = y) &= \sum_{y \in G} \sum_{x \in F} g(x,y) P(X = x \,|\, Y = y) P(Y = y) \\ &= \sum_{x} \sum_{y} g(x,y) P(X = x, Y = y) = E[g(X,Y)].\end{aligned}$$

In particular, if $E[g(X,Y)] < \infty$, $\sum_{y \in G} \psi(y) P(Y = y) < \infty$, which implies that (Theorem 2.1.27) $P(\psi(Y) < \infty) = 1$. Therefore, $E\left[E^Y\left[g(X,Y)\right]\right] < \infty$. Let now $g : F \times G \to \mathbb{R}$ be a function of arbitrary sign such that $E[|g(X,Y)|] < \infty$, and in particular $E[g^{\pm}(X,Y)] < \infty$. Denote by $\psi^{\pm}$ the functions associated with $g^{\pm}$ as in (2.21). As we just saw, for all $y \in G$, $\psi^{\pm}(y) < \infty$, and therefore $\psi(y) = \psi^{+}(y) - \psi^{-}(y)$ is well-defined (not an indeterminate $\infty - \infty$ form). Thus, the conditional expectation is well-defined also in the integrable case. From the observation made a few lines above, in this case,

$$|E^Y[g(X,Y)]| = |E^Y[g^{+}(X,Y)]| + |E^Y[g^{-}(X,Y)]| < \infty, \ P\text{-a.s.}$$

Definition 2.3.3 *The number $\psi(y)$ defined by (2.21) is called the conditional expectation of $g(X,Y)$ given $Y = y$, and is denoted by $E^{Y=y}[g(X,Y)]$ or, alternatively, by $E[g(X,Y) \,|\, Y = y]$. The random variable $\psi(Y)$ is called the conditional expectation of $g(X,Y)$ given Y, and is denoted by $E^Y[g(X,Y)]$ or $E[g(X,Y) \,|\, Y]$.*

Example 2.3.4: The hypergeometric distribution. Let X_1 and X_2 be independent binomial random variables of same size N and same parameter p. We are going to show that

$$E^{X_1+X_2}[X_1] = \psi(X_1 + X_2) = \frac{X_1 + X_2}{2}\,.$$

We have

$$\begin{aligned}P(X_1 = k | X_1 + X_2 = n) = {} & \frac{P(X_1 = k, X_1 + X_2 = n)}{P(X_1 + X_2 = n)} \\ & \frac{P(X_1 = k, X_2 = n - k)}{P(X_1 + X_2 = n)} \\ & \frac{P(X_1 = k) P(X_2 = n - k)}{P(X_1 + X_2 = n)}\,.\end{aligned}$$

Inserting the values of the probabilities thereof, and using the fact that the sum of two independent binomial random variables with size N and parameter p is a binomial random variable with size $2N$ and parameter p, a straightforward computation gives

$$P(X_1 = k | X_1 + X_2 = n) = \frac{\binom{N}{k}\binom{N}{n-k}}{\binom{2N}{n}}\,.$$

This is the hypergeometric distribution. The right-hand side of the last display is the probability of obtaining k black balls when a sample of n balls is randomly selected from an urn containing N black balls and N red balls. The mean of such a distribution is (by symmetry) $\frac{n}{2}$, therefore

$$E^{X_1+X_2=n}[X_1] = \frac{n}{2} = \psi(n)$$

and this gives the announced result. A more elegant solution is given in Exercise 2.4.22 where the reader will also discover that the result is more general.

Example 2.3.5: Two Poisson variables. Let X_1 and X_2 be two independent Poisson random variables with respective means $\theta_1 > 0$ and $\theta_2 > 0$. We seek to compute $E^{X_1+X_2}[X_1]$, that is $E^Y[X]$, where $X = X_1$, $Y = X_1 + X_2$. For $y \geq x$, the same computations as in Example 2.3.4 give

$$P(X = x \,|\, Y = y) = \frac{P(X_1 = x)P(X_2 = y - x)}{P(X_1 + X_2 = y)}.$$

Inserting the values of the the probabilities thereof, and using the fact that the sum of two independent Poisson random variables with parameter θ_1 and θ_2 is a Poisson random variable with parameter $\theta_1 + \theta_2$, a simple computation yields

$$P(X = x \,|\, Y = y) = \binom{y}{x}\left(\frac{\theta_1}{\theta_1+\theta_2}\right)^x\left(\frac{\theta_2}{\theta_1+\theta_2}\right)^{y-x}.$$

Therefore, with $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$,

$$\psi(y) = E^{Y=y}[X] = \sum_{x=0}^{y} x\binom{y}{x}\alpha^x(1-\alpha)^{y-x} = \alpha y\,.$$

Finally, $E^Y[X] = \psi(Y) = \alpha Y$, that is,

$$E^{X_1+X_2}[X_1] = \frac{\theta_1}{\theta_1+\theta_2}(X_1 + X_2)\,.$$

2.3.3 Basic Properties of Conditional Expectation

The first property of conditional expectation, *linearity*, is obvious from the definitions: For all $\lambda_1, \lambda_2 \in \mathbb{R}$,

$$E^Y[\lambda_1 g_1(X,Y) + \lambda_2 g_2(X,Y)] = \lambda_1 E^Y[g_1(X,Y)] + \lambda_2 E^Y[g_2(X,Y)]$$

whenever the conditional expectations thereof are well-defined and do not produce $\infty - \infty$ forms. *Monotonicity* is equally obvious: if $g_1 \leq g_2$, then

$$E^Y[g_1(X,Y)] \leq E^Y[g_2(X,Y)].$$

Theorem 2.3.6 *If g is non-negative or such that $E[|g(X,Y)|] < \infty$, we have*

$$E[E^Y[g(X,Y)]] = E[g(X,Y)].$$

Proof.

$$\begin{aligned} E[E^Y[g(X,Y)]] &= E[\psi(Y)]] = \sum_{y \in G} \psi(y) P(Y = y) \\ &= \sum_{y \in G} \sum_{x \in F} g(x,y) P(X = x \,|\, Y = y) P(Y = y) \\ &= \sum_{x} \sum_{y} g(x,y) P(X = x, Y = y) = E[g(X,Y)]. \end{aligned}$$

□

Theorem 2.3.7 *If w is non-negative or such that $E[|w(Y)|] < \infty$,*

$$E^Y[w(Y)] = w(Y),$$

and more generally,

$$E^Y[w(Y)h(X,Y)] = w(Y)E^Y[h(X,Y)]\,,$$

assuming that the left-hand side is well-defined.

Proof. We prove the second (more general) identity. We do this for non-negative w and h, the general case following easily from this special case:

$$\begin{aligned} E^{Y=y}[w(Y)h(X,Y)] &= \sum_{x \in F} w(y) h(x,y) P(X = x \,|\, Y = y) \\ &= w(y) \sum_{x \in F} h(x,y) P(X = x \,|\, Y = y) \\ &= w(y) E^{Y=y}[h(X,Y)]. \end{aligned}$$

□

Theorem 2.3.8 *If X and Y are independent and if v is non-negative or such that $E[|v(X)|] < \infty$, then*

$$E^Y[v(X)] = E[v(X)].$$

Proof. We have

$$\begin{aligned} E^{Y=y}[v(X)] &= \sum_{x \in F} v(x) P(X = x \,|\, Y = y) \\ &= \sum_{x \in F} v(x) P(X = x) = E[v(X)]. \end{aligned}$$

□

Theorem 2.3.9 *If X and Y are independent and if $g : F \times G \to \mathbb{R}$ is non-negative or such that $E[|g(X,Y)|] < \infty$, then, for all $y \in G$,*

$$E[g(X, Y \,|\, Y = y] = E[g(X, y)].$$

Proof. Applying formula (2.21) with $P(X = x \,|\, Y = y) = P(X = x)$ (by independence), we obtain

$$\psi(y) = \sum_{x \in F} g(x, y) P(X = x) = E\left[g(X, y)\right] .$$

□

Successive Conditioning

Suppose that $Y = (Y_1, Y_2)$, where Y_1 and Y_2 are discrete random variables. In this situation, we use the more developed notation

$$E^Y[g(X, Y)] = E^{Y_1, Y_2}[g(X, Y_1, Y_2] .$$

Theorem 2.3.10 *Let $Y = (Y_1, Y_2)$ be as above, and let $g : F \times G \to \mathbb{R}$ be either non-negative or such that $E[|g(X, Y)|] < \infty$. Then*

$$E^{Y_2}[E^{Y_1, Y_2}[g(X, Y_1, Y_2)]] = E^{Y_2}[g(X, Y_1, Y_2)].$$

Proof. Let

$$\psi(Y_1, Y_2) = E^{Y_1, Y_2}[g(X, Y_1, Y_2)].$$

We must show that

$$E^{Y_2}[\psi(Y_1, Y_2)] = E^{Y_2}[g(X, Y_1, Y_2)].$$

But

$$\psi(y_1, y_2) = \sum_x g(x, y_1, y_2) P(X = x \,|\, Y_1 = y_1, Y_2 = y_2)$$

and

$$E^{Y_2=y_2}[\psi(Y_1, Y_2)] = \sum_{y_1} \psi(y_1, y_2) P(Y_1 = y_1 \,|\, Y_2 = y_2),$$

that is,

$$\sum_{y_1} \sum_x g(x, y_1, y_2) P(X = x \,|\, Y_1 = y_1, Y_2 = y_2) P(Y_1 = y_1 \,|\, Y_2 = y_2).$$

But

$$\begin{aligned} P(X = x \,|\, Y_1 = y_1, Y_2 = y_2) & P(Y_1 = y_1 \,|\, Y_2 = y_2) \\ &= \frac{P(X = x,\, Y_1 = y_1, Y_2 = y_2)}{P(Y_1 = y_1, Y_2 = y_2)} \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} \\ &= P(X = x, Y_1 = y_1 \,|\, Y_2 = y_2)\,. \end{aligned}$$

Therefore

$$\begin{aligned} E^{Y_2=y_2}[\psi(Y_1, Y_2)] &= \sum_{y_1} \sum_x g(x, y_1, y_2) P(X = x, Y_1 = y_1 \,|\, Y_2 = y_2) \\ &= E^{Y_2=y_2}[g(X, Y_1, Y_2)]. \end{aligned}$$

□

Conditional Jensen's Inequality

Theorem 2.3.11 *Let I, φ and X be as in Theorem 3.1.5. Let Y be another random variable. Then*

$$E\left[\varphi(X) \,|\, Y\right] \geq \varphi(E\left[X \,|\, Y\right])\,.$$

Proof. The proof follows exactly the same lines as that of Theorem 3.1.5. □

The FKG Inequality

Theorem 2.3.12 *Let $E \subseteq \mathbb{R}$ and let $f, g : E^n \to \mathbb{R}$ be two bounded functions that are non-decreasing in each of their arguments. Let $X_1^n := (X_1, \ldots, X_n)$ be a vector of independent variables with values in E. Then,*

$$E\left[f(X_0^n) g(X_0^n)\right] \geq E\left[f(X_0^n)\right] E\left[g(X_0^n)\right]\,. \tag{2.22}$$

In other words, $f(X_0^n)$ and $g(X_0^n)$ are positively correlated.

Proof. By induction. For $n = 1$: Let X_1 and Y_1 be two independent and identically distributed E-valued random variables, and let $f, g : E \to \mathbb{R}_+$ be two non-decreasing bounded functions. Since $f(X_1) - f(Y_1)$ and $g(X_1) - g(Y_1)$ have the same sign, their product is non-negative, and therefore

$$E\left[(f(X_1) - f(Y_1))(g(X_1) - g(Y_1))\right] \geq 0\,.$$

Developing the left-hand side

$$E\left[f(X_1)g(X_1)\right] + E\left[f(Y_1)g(Y_1)\right] \geq E\left[f(X_1)\right] E\left[g(Y_1)\right] + E\left[f(Y_1)\right] E\left[g(X_1)\right]\,.$$

As X_1 and Y_1 have the same distribution, the left-hand side equals $2E\left[f(X_1)g(X_1)\right]$. Since X_1 and Y_1 have the same distribution and are independent, the right-hand side equals $2E\left[f(X_1)\right] E\left[g(X_1)\right]$. Therefore

$$E\left[f(X_1)g(X_1)\right] \geq E\left[f(X_1)\right] E\left[g(X_1)\right].$$

We now suppose that the result is true for $n - 1$ and show that it is then true for n. From the independence of X_0^{n-1} and X_n and Theorem 2.3.9,

$$E\left[f(X_0^n)g(X_0^n) \mid X_n = x_n\right] = E\left[f(X_0^{n-1}, x_n)g(X_0^{n-1}, x_n)\right]$$

and since, by the result assumed for $n - 1$,

$$\begin{aligned} E\left[f(X_0^{n-1}, x_n)g(X_0^{n-1}, x_n)\right] &\geq E\left[f(X_0^{n-1}, x_n)\right] E\left[g(X_0^{n-1}, x_n)\right] \\ &= E\left[f(X_0^n) \mid X_n = x_n\right] E\left[g(X_0^n) \mid X_n = x_n\right]\,, \end{aligned}$$

we have that

$$E\left[f(X_0^n)g(X_0^n) \mid X_n = x_n\right] \geq E\left[f(X_0^n) \mid X_n = x_n\right] E\left[g(X_0^n) \mid X_n = x_n\right]\,,$$

or

$$E\left[f(X_0^n)g(X_0^n) \mid X_n\right] \geq E\left[f(X_0^n) \mid X_n\right] E\left[g(X_0^n) \mid X_n\right]\,.$$

Taking expectations

$$\begin{aligned} E\left[f(X_0^n)g(X_0^n)\right] &\geq E\left[E\left[f(X_0^n) \mid X_n\right] E\left[g(X_0^n) \mid X_n\right]\right] \\ &\geq E\left[E\left[f(X_0^n) \mid X_n\right]\right] E\left[E\left[g(X_0^n) \mid X_n\right]\right] \\ &= E\left[f(X_0^n\right] E\left[g(X_0^n\right]\,, \end{aligned}$$

where the last inequality follows from the case $n = 1$ applied to the functions $x_n \to E\left[f(X_0^n) \mid X_n = x_n\right] = E\left[f(X_0^{n-1}, x_n)\right]$ and $x_n \to E\left[g(X_0^n) \mid X_n = x_n\right] = E\left[f(X_0^{n-1}, x_n)\right]$ which are non-decreasing. □

Remark 2.3.13 A stronger version of the above FKG inequality will be given in Section 9.3.

An Alternative Point of View

This subsection presents another definition of conditional expectation. It is the starting point for a generalization to the case of random elements that are not discrete. Even in the discrete case, this new perspective is indispensable (see Exercise 2.4.24).

Let X and Y be two discrete random variables with values in E and F respectively. Let $g : E \times F \to \mathbb{R}_+$ be a function that is either non-negative or such that $g(X,Y)$ is integrable. For any non-negative bounded function $\varphi : F \to \mathbb{R}$, we have

$$E\left[E^Y\left[g(X,Y)\right]\varphi(Y)\right] = E\left[g(X,Y)\varphi(Y)\right] . \qquad (\star)$$

In fact,

$$\begin{aligned} E\left[E^Y\left[g(X,Y)\right]\varphi(Y)\right] &= E\left[\psi(Y)\varphi(Y)\right] = \sum_{y\in F}\psi(y)\varphi(y)P(Y=y) \\ &= \sum_{y\in F}\left(\sum_{x\in E} g(x,y)\frac{P(X=x,Y=y)}{P(Y=y)}\,dx\right)\varphi(y)P(Y=y) \\ &= \sum_{y\in F}\sum_{x\in E} g(x,y)\varphi(y)P(X=x,Y=y) = E\left[g(X,Y)\varphi(Y)\right] . \end{aligned}$$

This suggests to take $(\star)$ as a basis for an extension of the definition of conditional expectation. The conditioned variable is now any random element Z taking its values in E, a denumerable subset of $\mathbb{R}$.

Definition 2.3.14 *Let Z and Y be as above, and suppose that Z is either non-negative or integrable. A conditional expectation $E^Y\left[Z\right]$ is by definition a random variable of the form $\psi(Y)$ such that equality*

$$E\left[\psi(Y)\varphi(Y)\right] = E\left[Z\varphi(Y)\right] \qquad (2.23)$$

holds for any non-negative bounded function $\varphi : E \to \mathbb{R}$.

Theorem 2.3.15 *In the situation described in the above definition, the conditional expectation exists and is essentially unique.*

By "essentially unique" the following is meant. If there are two functions ψ_1 and ψ_2 that meet the requirement, then $\psi_1(Y) = \psi_2(Y)$ almost surely.

Proof. The proof of existence is by the construction at the begining of the section, replacing $g(X,Y)$ by Z (more explicitly, $h : E \to \mathbb{R}$, $X = Z$, $g(x,y) = h(z)$). For uniqueness, suppose that ψ_1 and ψ_2 meet the requirement. In particular $E\left[\psi_1(Y)\varphi(Y)\right] = E\left[\psi_2(Y)\varphi(Y)\right]$ $(= E\left[Z\varphi(Y)\right])$, or $E\left[(\psi_1(Y) - \psi_2(Y))\varphi(Y)\right] = 0$, for any non-negative bounded function $\varphi : \mathbb{R}^n \to \mathbb{R}$. Choose $\varphi(Y) = 1_{\{\psi_1(Y)-\psi_2(Y)>0\}}$ to obtain

$$E\left[(\psi_1(Y)-\psi_2(Y))1_{\{\psi_1(Y)-\psi_2(Y)>0\}}\right]=0\,.$$

Since the random variable $(\psi_1(Y)-\psi_2(Y))1_{\{\psi_1(Y)-\psi_2(Y)>0\}}$ is non-negative and has a null expectation, it must be almost surely null. In other terms $\psi_1(Y)-\psi_2(Y)\le 0$ almost surely. Exchanging the roles of ψ_1 and ψ_2, we have that $\psi_1(Y)-\psi_2(Y)\ge 0$ almost surely. Therefore $\psi_1(Y)-\psi_2(Y)=0$ almost surely. □

Example 2.3.16: Let Y be a positive integer-valued random variable.

$$E^Y[Z]=\sum_{n=1}^{\infty}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}1_{\{Y=n\}},$$

where, by convention, $\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}=0$ when $P(Y=n)=0$ (in other terms, the sum in the above display is over all n such that $P(Y=n)>0$).

Proof. We must verify (2.23) for all bounded measurable $\varphi:\mathbb{R}\to\mathbb{R}$. The right-hand side is equal to

$$\begin{aligned}
&E\left[\left(\sum_{n\ge 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}1_{\{Y=n\}}\right)\left(\sum_{k\ge 1}\varphi(k)1_{\{Y=k\}}\right)\right]\\
&=E\left[\sum_{n\ge 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}\varphi(n)1_{\{Y=n\}}\right]=\sum_{n\ge 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}\varphi(n)E[1_{\{Y=n\}}]\\
&=\sum_{n\ge 1}\frac{E[Z1_{\{Y=n\}}]}{P(Y=n)}\varphi(n)P(Y=n)=\sum_{n\ge 1}E[Z1_{\{Y=n\}}]\varphi(n)\\
&=\sum_{n\ge 1}E[Z1_{\{Y=n\}}\varphi(n)]=E[Z(\sum_{n\ge 1}\varphi(n)1_{\{Y=n\}})]=E[Z\varphi(Y)]\,.
\end{aligned}$$

□

2.4 Exercises

Exercise 2.4.1. Geometric

Let T_1 and T_2 be two independent geometric random variables with the same parameter $p\in(0,1)$. Give the probability distribution of their sum $X=T_1+T_2$.

Exercise 2.4.2. Variance of the coupon's collector variable

In the coupon's collector problem of Example 2.1.42, compute the variance σ_X^2 of X (the number of chocolate tablets needed to complete the collection of the n different coupons) and show that $\frac{\sigma_X^2}{n^2}$ has a limit (to be identified) as $n\uparrow\infty$.

Exercise 2.4.3. Poisson

1. Let X be a Poisson random variable with mean $\theta>0$. Compute the mean of the random variable $X!$ (factorial, not exclamation mark).

2. Compute $E\left[\theta^X\right]$.

3. What is the probability that X is odd?

Exercise 2.4.4. Random sum
Let $\{X_n\}_{n\geq 1}$ be independent random variables taking the values 0 and 1 with probability $q = 1-p$ and p, respectively, where $p \in (0,1)$. Let T be a Poisson random variable with mean $\theta > 0$, independent of $\{X_n\}_{n\geq 1}$. Compute the probability distribution of $S := X_1 + \cdots + X_T$.

Exercise 2.4.5. The binomial random variable
(a) Let $X \sim \mathcal{B}(n,p)$. Show that $Y := n - X \sim \mathcal{B}(n, 1-p)$.
(b) Let $X_1, \ldots, X_{2n}$ be independent random variables taking the values 0 or 1, and such that for all i, $P(X_i = 1) = p \in (0,1)$. Give the probability distribution of the random variable $Z := \sum_{i=1}^{n} X_i \, X_{n+i}$.

Exercise 2.4.6. Null variance
Let X be a discrete random variable taking its values in E, with probability distribution $p(x)$, $x \in E$.
(i) Let $A := \{\omega;\, p(X(\omega)) = 0\}$. Show that $P(A) = 0$.
(ii) Prove that a real-valued random variable with null variance is almost surely constant.

Exercise 2.4.7. The blue pinko
The blue pinko is a bird owing its name to the fact that it lays eggs that are either blue or pink. Suppose that it lays T eggs, with probability p that a given egg is blue, and that the colours of the successive eggs are independent and independent of the total number of eggs. The conclusion of Exercise 2.4.4 was that if the number of eggs is Poisson with mean θ, then the number of blue eggs is a Poisson random variable with mean θp and the number of pink eggs is a Poisson random variable with mean $\theta(1-p)$. Prove that the number of blue eggs and the number of pink eggs are independent random variables.

Exercise 2.4.8. The entomologist
Each individual of a specific breed of insects has, independently of the others, the probability θ of being a male.

(A) An entomologist seeks to collect exactly $M > 1$ males, and therefore stops hunting as soon as she captures M males. What is the distribution of X, the number of insects she must catch to collect *exactly* M males?

(B) What is the distribution of X, the smallest number of insects that the entomologist must catch to collect *at least* M males and N females?

Exercise 2.4.9. Maximal bin load
N balls are thrown independently and at random in N bins. This results in X_i balls in bin i $(1 \leq i \leq N)$. Let $X_{max} = \max\{X_1, \ldots, X_N\}$ be the maximal bin load. Prove the following: For sufficiently large N,

$$P\left(X_{max} > \frac{\log N}{\log^{(2)} N}\right) \geq 1 - \frac{1}{N},$$

where $\log^{(2)} N := \log(\log N)$.

Exercise 2.4.10. The matchbox
A smoker has one matchbox with n matches in each pocket. He reaches at random for one box or the other. What is the probability that, having eventually found an empty matchbox, there will be k matches left in the other box?

Exercise 2.4.11. Biased dice and uniformity
Is it possible to have two biased dice such that tossing them independently results in a total number of points uniformly distributed on $\{2, 3, \dots, 12\}$?

Exercise 2.4.12. Residual time
Let X be a random variable with values in $\mathbb{N}$ and with finite mean m. Show that $p_n = \frac{1}{m} P(X > n)$ $(n \geq 0)$ defines a probability distribution on $\mathbb{N}$ and compute its generating function in terms of the generating function of X.

Exercise 2.4.13. Mean and variance via generating functions
(a) Compute the mean and variance of the binomial random variable B of size n and parameter p from its generating function. Do the same for the Poisson random variable P of mean θ.

(b) What is the generating function g_T of the geometric random variable T with parameter $p \in (0, 1)$? Compute its first two derivatives and deduce from the result the variance of T.

(c) What is the n-th factorial moment $(E\left[X(X-1)\cdots(X-n+1)\right])$ of a Poisson random variable X of mean $\theta > 0$?

Exercise 2.4.14. From generating function to distribution
What is the probability distribution of the integer-valued random variable with generating function $g(z) = \frac{1}{(2-z)^2}$? Compute the fifth moment $(E[X^5])$ of this random variable.

Exercise 2.4.15. Throw a die
You perform three independent tosses of an unbiased die. What is the probability that one of these tosses results in a number that is the sum of the two other numbers? (You are required to find a solution using generating functions.)

Exercise 2.4.16. Generalized Wald's formula
Let $\{Y_n\}_{n\geq 1}$ be a sequence of integer-valued integrable random variables such that $E[Y_n] = E[Y_1]$ for all $n \geq 1$. Let T be an integer-valued random variable such that for all $n \geq 1$, the event $\{T \geq n\}$ is independent of Y_n. Let $X := \sum_{n=1}^{T} Y_n$. Prove that $E\left[X\right] = E[Y_1]E[T]$.

Exercise 2.4.17. When Wald's formula does not apply
Let $\{Y_n\}_{n\geq 1}$ be a sequence of integer-valued integrable random variables such that $E[Y_n] = E[Y_1]$ for all $n \geq 1$. Let T be an integer-valued random variable. Let $X := \sum_{n=1}^{T} Y_n$. It is not true in general that $E[X] = E[Y_1]E[T]$. Give a simple counterexample.

Exercise 2.4.18. The return of the entomologist
Recall the setup of Exercise 2.4.8. What is the expectation of X, the number of insects the entomologist must capture to collect *exactly* M males? (In Exercise 2.4.8, you computed the distribution of X, from which you can of course compute the mean. However, you can give the solution directly, and this is what is required in the present exercise.)

Exercise 2.4.19. Conditioning by sampling
Let Z be a discrete random variable with values in E and let $f : E \to \mathbb{R}$ be a non-negative function. Let $\{Z_n\}_{n\geq 1}$ be an IID sequence of random variables with values in E and the same distribution as Z. Let A be some subset of E such that $P(Z \in A) > 0$.

(1) Define the random variable τ to be the first time $n \geq 1$ such that $Z_n \in A$. Prove that $P(\tau < \infty) = 1$.

(2) Let Z_τ be the random variable equal to Z_n when $\tau = n$. Prove that

$$E[f(Z_\tau)] = E[f(Z) \mid Z \in A] .$$

Exercise 2.4.20. Multinomial distribution and conditioning
Let $(X_1, \dots, X_k)$ be a multinomial random vector with size n and parameters $p_1, \dots, p_k$. Compute $E^{X_1}[X_2 + \cdots + X_{k-1}]$ and $E^{X_1}[X_2]$.

Exercise 2.4.21. XYZ
Let X, Y, and Z be three discrete random variables with values in E, F, and G, respectively. Prove the following: If for some function $g : E \times F \to [0,1]$, $P(X = x \,|\, Y = y, Z = z) = g(x,y)$ for all x, y, z, then $P(X = x \,|\, Y = y) = g(x,y)$ for all x, y, and X and Z are conditionally independent given Y.

Exercise 2.4.22. A natural result
Let X_1 and X_2 be two integrable independent identically distributed discrete real-valued random variables. Prove that

$$E^{X_1+X_2}[X_1] = \frac{X_1 + X_2}{2}.$$

Exercise 2.4.23. Pólya's urn
There is an urn containing black balls and white balls, the number of which varies in time as follows. At time $n = 0$ there is one black ball and one white ball. At a given time one of the balls is selected at random, its colour is observed, and the ball is replaced in the urn together with a new ball of the same colour. In particular

the number of balls increases by one unit at each draw. Let B_k be the number of black balls after exactly k balls have been added. Prove that B_k is uniformly distributed on $\{1, 2, \dots, k+1\}$.

Exercise 2.4.24. Conditioning by the square
Let X be a random variable with values in $\mathbb{Z}$ and probability distribution $(p(n)\,,\,n \in \mathbb{Z})$. Let $h : \mathbb{Z} \to \mathbb{R}$ be a function such that $E\left[|h(Z)|\right] < \infty$. Prove formally that

$$E\left[h(X)\,|\,X^2\right] = h(|X|)\frac{p(|X|)}{p(|X|) + p(-|X|)} + h(-|X|)\frac{p(-|X|)}{p(|X|) + p(-|X|)}\,.$$

Exercise 2.4.25. Bayesian tests of hypotheses
Let Θ be a discrete random variable with values in $\{1, 2, ..., K\}$ and let X be a discrete random variable with values in E. The joint distribution of Θ and X is specified in the following manner. For all $1 \leq i \leq K$,

$$P\left(\Theta = i\right) = \pi(i), \qquad P\left(X = x|\Theta = i\right) = p_i(x),$$

where π is a probability distribution on $\{1, 2, ..., K\}$ and the p_i's are probability disributions on E.

These random variables may be interpreted in terms of tests of hypotheses. The variable Θ represents the state of Nature, and X — called the observation — is the (random) result of an experiment that depends on the actual state of Nature. If Nature happens to be in state i, then X admits the distribution p_i.

In view of the observation X, we wish to infer the actual value of Θ. For this, we design a guess strategy, that is a function $g : E \to \{1, 2, ..., K\}$ with the interpretation that $\widehat{\Theta} := g(X)$ is our guess (based only on the observation X) of the (not directly observed) state Θ of Nature. An equivalent description of the strategy g is the partition $\mathcal{A} = \{A_1, \dots, A_K\}$ of $\mathbb{R}^m$ given by $A_i := \{x \in E;\, g(x) = i\}$. The decision rule is then

$$X \in A_i \Rightarrow \widehat{\Theta} = i\,.$$

Prove the following: Any partition $\mathcal{A}^*$ such that

$$x \in A_i^* \Rightarrow \pi(i)p_i(x) = \max_k \left(\pi(k)p_k(x)\right)$$

minimizes the probability of error P_E.

http://www.springer.com/978-3-319-43475-9

Discrete Probability Models and Methods
Probability on Graphs and Trees, Markov Chains and
Random Fields, Entropy and Coding
Brémaud, P.
2017, XIV, 559 p. 92 illus., Hardcover
ISBN: 978-3-319-43475-9