3 Discrete Random Variables

(3.1) Definition. Let Ω be a sample space and $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ a set of events with a probability measure $\Pr : \mathcal{A} \to \mathbb{R}$. Given $(\Omega, \mathcal{A}, \Pr)$, a mapping $X : \Omega \to \mathbb{R}$ is called a *discrete random variable* if

$$X^{-1}(x) \ \in \ \mathcal{A} \qquad \text{for all } \ x \in \mathbb{R}$$

and there exist countable (finite or infinite) many distinct real numbers $(x_i)_{i \in I}$ such that:

$$\sum_{i \in I} \Pr(X^{-1}(x_i)) \ = \ 1$$

If X is a discrete random variable, its *probability mass function* (pmf) is defined by:

$$p_X : \mathbb{R} \to [0,1], \qquad p_X(x) \ := \ \Pr(X = x) \ := \ \Pr(X^{-1}(x)) \quad \text{for all } \ x \in \mathbb{R}$$

(3.2) Lemma. Let $(\Omega, \mathcal{A}, \Pr)$ be as in (3.1) and $X : \Omega \to \mathbb{R}$ a discrete random variable. Then, for any subset $S \subseteq \mathbb{R}$ with $X^{-1}(S) \in \mathcal{A}$:

$$\Pr(X \in S) \ := \ \Pr(X^{-1}(S)) \ = \ \Pr(\{\omega \in \Omega \mid X(\omega) \in S\}) \ = \ \sum_{\substack{x \in S \\ p_X(x) \neq 0}} p_X(x)$$

Proof: Let

$$P \ := \ \{x \in \mathbb{R} \mid p_X(x) > 0\}$$

and:

$$N \ := \ \Omega \setminus X^{-1}(P)$$

Then

$$\Omega \ = \ N \ \dot\cup \ \dot{\bigcup_{x \in P}} X^{-1}(x)$$

and

$$1 \ = \ \Pr(\Omega) \ = \ \Pr(N) \ + \sum_{x \in P} \Pr\left(X^{-1}(x)\right) \ = \ \Pr(N) \ + \ 1$$

shows that

$$\Pr(N) \ = \ 0$$

and:

$$\begin{aligned}
\Pr(X^{-1}(S)) \ &= \ \Pr(\{\omega \in \Omega \mid X(\omega) \in S\}) \\
&= \ \Pr\left((X^{-1}(S) \cap N) \ \dot\cup \ \dot{\bigcup_{x \in S \cap P}} X^{-1}(x) \right) \\
&= \ \Pr\left(X^{-1}(S) \cap N\right) \ + \sum_{x \in S \cap P} \Pr\left(X^{-1}(x)\right) \\
&= \ \sum_{x \in S \cap P} \Pr\left(X^{-1}(x)\right) \ = \ \sum_{\substack{x \in S \\ p_X(x) \neq 0}} p_X(x)
\end{aligned}$$

(3.3) Notation. If $X : \Omega \to \mathbb{R}$ is a discrete random variable, then for every subset $S \subseteq \mathbb{R}$

$$\sum_{x \in S} p_X(x) \quad := \quad \sum_{\substack{x \in S \\ p_X(x) \neq 0}} p_X(x)$$

is well defined and will be denoted by $\Pr(X \in S)$, even if $X^{-1}(S) \notin \mathcal{A}$.

(3.4) Remark. Even if $X^{-1}(S) \notin \mathcal{A}$, "the event" $X^{-1}(S)$ may be defined. It is the smallest set in $\mathcal{A}$ containing $X^{-1}(S)$, i.e.:

$$\left(X^{-1}(S)\right)_{\mathcal{A}} \quad := \quad \bigcap_{\substack{A \in \mathcal{A} \\ X^{-1}(S) \subseteq A}} A$$

Note: If N is defined as in the proof of (3.2), then

$$(X^{-1}(S))_{\mathcal{A}} \;\subseteq\; X^{-1}(S) \cup N \;\in\; \mathcal{A}$$

and:

$$\Pr(X \in S) \;=\; \Pr\left(\left(X^{-1}(S)\right)_{\mathcal{A}}\right) \;=\; \Pr\left(X^{-1}(S) \cup N\right)$$

(3.5) Notation. In line with the introduced notations $\Pr(X = x)$ and $\Pr(X \in S)$, the probability that $X(\omega)$ belongs to some interval $S = (a, b]$ is denoted by:

$$\Pr(a < X \leq b) \quad := \quad \Pr(X \in (a, b])$$

Similar notations are used for other types of intervals.

(3.6) Definition. The *cumulative distribution function* (cdf) of a random variable X is defined by:

$$F_X : \mathbb{R} \to [0, 1]\,, \qquad F_X(x) \;=\; \Pr(X \leq x)$$

(3.7) Discrete sample spaces of real numbers. If Ω is a countable subset of $\mathbb{R}$, i.e. $\Omega = \{\omega_i | i \in I\} \subseteq \mathbb{R}$ for some countable set I, and Pr is a probability on $\mathcal{A} = \mathcal{P}(\Omega)$ (i.e. $p_i = \Pr(\omega_i)$ is known for all $i \in I$), then the embedding $X : \Omega \to \mathbb{R}$ with $X(\omega) = \omega$ for all $\omega \in \Omega$ is a random variable. By virtue of such embeddings, all notions related to random variables (like expectation, variance, etc., which will be defined later) can also be applied to discrete sample spaces of real numbers.

(3.8) Remark. Any function $p : \mathbb{R} \to [0, 1]$ with countable support $\Omega := \{\omega \in \mathbb{R} \mid p(\omega) \neq 0\}$ and $\sum_{\omega \in \Omega} p(\omega) = 1$ is a probability mass function $p = p_X$, defined by:

- sample space $\Omega := \{\omega \in \mathbb{R} \mid p(\omega) \neq 0\}$,
- probability $\Pr : \mathcal{P}(\Omega) \to [0, 1]$ with $\Pr(\{\omega\}) = p(\omega)$ for all $\omega \in \Omega$, and
- random variable $X : \Omega \hookrightarrow \mathbb{R}$ with $X(\omega) = \omega$ for all $\omega \in \Omega$

(3.9) Example. Consider the sample space

$$\Omega \quad = \quad \{(i, j) \mid i, j \in \{1, 2, 3, 4, 5, 6\}\}$$

with a uniform probability measure, i.e.

$$\Pr(\{\omega\}) \;=\; \frac{1}{36}$$

for every $\omega \in \Omega$.

If $X : \Omega \to \mathbb{R}$ is defined to be the sum of the two components of ω, i.e. $X((i,j)) = i + j$, its pmf is as follows:

$$\begin{aligned} p_X(2) &= p_X(12) = \tfrac{1}{36}, \quad p_X(3) = p_X(11) = \tfrac{2}{36}, \quad p_X(4) = p_X(10) = \tfrac{3}{36}, \\ p_X(5) &= p_X(9) = \tfrac{4}{36}, \quad p_X(6) = p_X(8) = \tfrac{5}{36}, \quad p_X(7) = \tfrac{6}{36} \end{aligned}$$

(3.10) Definition. Let X be discrete random variable with pmf $p_X(x)$. The *expectation* (*mean*) of X is defined to be

$$E(X) \;:=\; \sum_{x\in\mathbb{R}} x \cdot p_X(x)$$

if at least one of the sums $\sum\limits_{x\in\mathbb{R}^-} x \cdot p_X(x)$ or $\sum\limits_{x\in\mathbb{R}^+} x \cdot p_X(x)$ is finite.

(3.11) Lemma. Let $X : \Omega \to \mathbb{R}$ be a random variable with countable sample space Ω and $\mathcal{A} = \mathcal{P}(\Omega)$. Then:

$$E(X) \;=\; \sum_{\omega\in\Omega} X(\omega) \cdot \Pr(\omega)$$

(3.12) Lemma. Let X be a discrete random variable. Then, every function $g : \mathbb{R} \to \mathbb{R}$ defines a random variable:

$$Y \;=\; g(X) \;:=\; g \circ X$$

(i) The pmf of Y is given by:

$$p_Y(y) \;=\; \sum_{x\in g^{-1}(y)} p_X(x) \qquad \text{for all } \; y \in \mathbb{R}$$

(ii) If $E(Y)$ exists, then:

$$E(Y) \;=\; \sum_{x\in\mathbb{R}} g(x) \cdot p_X(x)$$

Proof:

(i) For any $y \in \mathbb{R}$:

$$\begin{aligned} p_Y(y) &= \Pr(Y = y) \\ &= \Pr(\{\omega | g(X(\omega)) = y\}) \\ &= \Pr(\{\omega | X(\omega) \in g^{-1}(y)\}) \\ &= \sum_{x \in g^{-1}(y)} \Pr(\{\omega | X(\omega) = x\}) \\ &= \sum_{x \in g^{-1}(y)} p_X(x) \end{aligned}$$

(ii) As

$$\mathbb{R} = \dot{\bigcup_{y \in \mathbb{R}}} g^{-1}(y) \qquad (*)$$

there are only countable many different $y \in \mathbb{R}$ with $p_Y(y) \neq 0$, and:

$$\begin{aligned} E(Y) &= \sum_{y \in \mathbb{R}} y \cdot p_Y(y) \\ &= \sum_{y \in \mathbb{R}} \left(y \sum_{x \in g^{-1}(y)} p_X(x) \right) \\ &= \sum_{y \in \mathbb{R}} \sum_{x \in g^{-1}(y)} y \, p_X(x) \\ &= \sum_{y \in \mathbb{R}} \sum_{x \in g^{-1}(y)} g(x) \, p_X(x) \\ &\overset{(*)}{=} \sum_{x \in \mathbb{R}} g(x) \, p_X(x) \end{aligned}$$

□

(3.13) Lemma. Let X be a discrete random variable, such that $E(X)$ exists. Then, for any $a, b \in \mathbb{R}$:

$$E(a\,X + b) = a\,E(X) + b$$

Proof: Follows from (3.12)(ii) with $g(x) := ax + b$. □

(3.14) Definition. Let X be a discrete random variable with expectation $m = E(X)$. The *variance* of X is defined to be:

$$\sigma^2 := \mathrm{Var}(X) := E((X - m)^2)$$

The *standard deviation* of X is defined to be:

$$\sigma := \sqrt{\mathrm{Var}(X)}$$

(3.15) Lemma.

$$\mathrm{Var}(X) \;=\; E(X^2) - (E(X))^2$$

Proof: Applying (3.12)(i) to $g_1 : x \mapsto (x-m)^2$ and $g_2 : x \mapsto x^2$ yields:

$$\begin{aligned}
E((X-m)^2) &= \sum_{x\in\mathbb{R}} (x^2 - 2mx + m^2)\, p_X(x) \\
&= \sum_{x\in\mathbb{R}} x^2 p_X(x) \;-\; 2m \sum_{x\in\mathbb{R}} x p_X(x) \;+\; m^2 \sum_{x\in\mathbb{R}} p_X(x) \\
&= E(X^2) \;-\; 2mE(X) \;+\; m^2 \\
&= E(X^2) \;-\; m^2
\end{aligned}$$

□

(3.16) Definition. For $n \in \mathbb{N}$, the n-th moment of a discrete random variable X is defined to be $E(X^n)$.

(3.17) Definition. The moment generating function $\phi_X(t)$ of a discrete random variable X with pmf $p_X(x)$ is defined by:

$$\phi_X(t) \;:=\; E\left(e^{tX}\right) \;=\; \sum_{x\in\mathbb{R}} e^{tx} \cdot p_X(x)$$

(3.18) Lemma. Suppose the moment generating function $\phi_X(t)$ is finite in an open interval around 0. Then all moments of X exist and are given by:

$$\phi_X^{(n)}(0) \;=\; E(X^n)$$

Proof:

$$\begin{aligned}
\phi_X(t) \;=\; E(e^{tX}) &= \sum_{x\in\mathbb{R}} e^{tx} \cdot p_X(x) \\
&= \sum_{x\in\mathbb{R}} \left(1 \;+\; tx \;+\; \frac{(tx)^2}{2!} \;+\; \frac{(tx)^3}{3!} + \ldots\right) \cdot p_X(x) \\
&= \sum_{x\in\mathbb{R}} p_X(x) \;+\; \left(\sum_{x\in\mathbb{R}} x\, p_X(x)\right) t \;+\; \frac{1}{2!}\left(\sum_{x\in\mathbb{R}} x^2\, p_X(x)\right) t^2 \;+\; \ldots \\
&= 1 \;+\; E(X)t \;+\; \frac{E(X^2)}{2!} t^2 \;+\; \frac{E(X^3)}{3!} t^3 \;+\; \ldots
\end{aligned}$$

On the other hand

$$\phi_X(t) \;=\; \phi_X(0) \;+\; \phi'_X(0)t \;+\; \frac{\phi_X^{(2)}(0)}{2!} t^2 \;+\; \frac{\phi_X^{(3)}(0)}{3!} t^3 \;+\; \ldots$$

and the claim follows from a comparison of the coefficients of the two power series representations of $\phi_X(t)$. □

(3.19) Definition. The *characteristic function* $\varphi_X(v)$ of a discrete random variable X with pmf $p_X(x)$ is defined by:

$$\varphi_X(v) \;:=\; E\left(e^{(j\cdot v)X}\right) \;=\; \sum_{x\in\mathbb{R}} e^{(j\cdot v)x}\cdot p_X(x)$$

(3.20) Remark. If $p_X(x)=0$ for all $x\in\mathbb{R}\setminus\mathbb{Z}$, then

$$\varphi_X(v) \;=\; \sum_{k\in\mathbb{Z}} e^{(j\cdot v)k}\cdot p_X(k)$$

is a 2π-periodic Fourier series and for all $k\in\mathbb{Z}$:

$$p_X(k) \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-jvk}\varphi_X(v)\,dv$$

Joint probabilities

(3.21) Definition. Let X, Y be discrete random variables with respect to the same probability measure Pr, i.e. with respect to the same triple $(\Omega, \mathcal{A}, \Pr)$. The *joint probability mass function* of X and Y, $p_{XY} : \mathbb{R}^2 \to [0,1]$, is defined by:

$$p_{XY}(x,y) \;:=\; \Pr(X=x, Y=y) \;=\; \Pr\left(X^{-1}(x) \cap Y^{-1}(y)\right) \qquad \text{for all } \; x, y \in \mathbb{R}$$

(3.22) Lemma. Let p_{XY} be the joint probability mass function of two random variables X and Y. The probability mass functions of X and Y are determined from the *marginal probability mass functions* of p_{XY} as follows:

$$p_X(x) \;=\; \sum_{y \in \mathbb{R}} p_{XY}(x,y)$$

$$p_Y(y) \;=\; \sum_{x \in \mathbb{R}} p_{XY}(x,y)$$

(3.23) Remark. Given two random variables whose probability mass functions are $\neq 0$ for only finitely many numbers, i.e. there are $m, n \in \mathbb{N}$ such that

$$\mathcal{X} \;:=\; \{x \in \mathbb{R} \mid p_X(x) \neq 0\} \;=\; \{x_1, \ldots, x_m\}$$

and:

$$\mathcal{Y} \;:=\; \{y \in \mathbb{R} \mid p_Y(y) \neq 0\} \;=\; \{y_1, \ldots, y_n\}$$

If $p_{XY}(x,y) \neq 0$ for some $(x,y) \in \mathbb{R}^2$, then $p_X(x) \neq 0$ and $p_Y(y) \neq 0$ by (3.22), i.e. $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Hence $p_{XY}(x,y) \neq 0$ may only hold, if $(x,y) \in \mathcal{X} \times \mathcal{Y}$, and p_{XY} is completely determined by the $m \cdot n$ numbers:

$$p_{ij} \;:=\; p_{XY}(x_i, y_j) \qquad (i = 1, \ldots, m, \; j = 1, \ldots, n)$$

Putting these values in a tabular scheme, the values for the marginal probability mass functions can be calculated by summing up all entries from a row or column of the table, respectively:

	y_1	y_2	$\cdots$	y_n	
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\sum_{j=1}^n p_{1j} = p_X(x_1)$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\sum_{j=1}^n p_{2j} = p_X(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	$\sum_{j=1}^n p_{mj} = p_X(x_m)$
	$\sum_{i=1}^m p_{i1}$ $= p_Y(y_1)$	$\sum_{i=1}^m p_{i2}$ $= p_Y(y_2)$	$\cdots$	$\sum_{i=1}^m p_{in}$ $= p_Y(y_n)$	

(3.24) Lemma. Let p_{XY} be the joint probability mass function of two random variables X, Y. Given any function $g : \mathbb{R}^2 \to \mathbb{R}$, a random variable

$$Z \;=\; g(X,Y) : \Omega \to \mathbb{R}$$

can be defined by

$$Z(\omega) \;=\; g(X,Y)(\omega) \;=\; g(X(\omega), Y(\omega)) \qquad \text{for all } \; \omega \in \Omega$$

and the following holds:

(i)
$$p_Z(z) \;=\; \sum_{(x,y)\in g^{-1}(z)} p_{XY}(x,y) \qquad \text{for all } \; z \in \mathbb{R}$$

(ii)
$$E(Z) \;=\; \sum_{(x,y)\in\mathbb{R}^2} g(x,y) \cdot p_{XY}(x,y)$$

(3.25) Lemma. Let X, Y be discrete random variables with respect to the same probability measure Pr, i.e. with respect to the same triple $(\Omega, \mathcal{A}, \Pr)$. Then:

$$E(X+Y) \;=\; E(X) \;+\; E(Y)$$

Proof:

$$\begin{aligned}
E(X+Y) &\overset{(3.24)(\text{ii})}{=} \sum_{(x,y)\in\mathbb{R}^2} (x+y) \cdot p_{XY}(x,y) \\
&= \sum_{(x,y)\in\mathbb{R}^2} x \cdot p_{XY}(x,y) \;+\; \sum_{(x,y)\in\mathbb{R}^2} y \cdot p_{XY}(x,y) \\
&= \sum_{x\in\mathbb{R}} x \sum_{y\in\mathbb{R}} p_{XY}(x,y) \;+\; \sum_{y\in\mathbb{R}} y \sum_{x\in\mathbb{R}} p_{XY}(x,y) \\
&= \sum_{x\in\mathbb{R}} x \cdot p_X(x) \;+\; \sum_{y\in\mathbb{R}} y \cdot p_Y(y) \\
&= E(X) \;+\; E(Y)
\end{aligned}$$

□

(3.26) Definition. Let X, Y be discrete random variables with respect to the same probability measure Pr, i.e. with respect to the same triple $(\Omega, \mathcal{A}, \Pr)$. Then X and Y are called *independent random variables* if for all subsets $S, T \subseteq \mathbb{R}$ the events $X^{-1}(S)$ and $Y^{-1}(T)$ are independent, i.e.:

$$\Pr(X \in S, Y \in T) \;=\; \Pr(X \in S) \cdot \Pr(Y \in T)$$

More generally, any finite number of random variables $X_1, \ldots, X_n$ are *independent*, if

$$\Pr\left(\bigcap_{i=1}^{n} \{X_i \in S_i\}\right) \;=\; \prod_{i=1}^{n} \Pr(X_i \in S_i)$$

for all subsets $S_1, \ldots, S_n \subseteq \mathbb{R}$.

(3.27) Lemma. Let X and Y be independent random variables and $h, k : \mathbb{R} \to \mathbb{R}$. Then also $h(X)$ and $k(Y)$ are independent random variables.

Proof: Let $S, T \subseteq \mathbb{R}$. Then the events $(h(X))^{-1}(S)$ and $(k(Y))^{-1}(T)$ are independent, because

$$(h(X))^{-1}(S) \;=\; X^{-1}\Big(h^{-1}(S)\Big), \quad (k(Y))^{-1}(T) \;=\; Y^{-1}\Big(k^{-1}(T)\Big)$$

and X and Y are independent random variables.

(3.28) Lemma. Let X, Y be discrete random variables with respect to the same probability measure Pr, i.e. with respect to the same triple $(\Omega, \mathcal{A}, \Pr)$. Then the following conditions are equivalent:

(i) X and Y are independent

(ii) $p_{XY}(x, y) \;=\; p_X(x) \cdot p_Y(y)$ for all $x, y \in \mathbb{R}$

(iii) $E(h(X) \cdot k(Y)) \;=\; E(h(X)) \cdot E(k(Y))$ for all functions $h, k : \mathbb{R} \to \mathbb{R}$

Proof: In order to prove the lemma, it suffices to show:

$$\text{(i)} \implies \text{(ii)} \implies \text{(iii)} \implies \text{(i)}$$

(i) $\implies$ (ii): Trivial, as (ii) simply states the independence of events $X^{-1}(S)$ and $Y^{-1}(T)$, where S and T just consist of single elements: $S = \{x\}$, $T = \{y\}$.

(ii) $\implies$ (iii):

$$\begin{aligned}
E(h(X) \cdot k(Y)) &= \sum_{(x,y)\in\mathbb{R}^2} h(x)\, k(y) \cdot p_{XY}(x, y) && \text{(by (3.24))} \\
&= \sum_{x\in\mathbb{R}} \sum_{y\in\mathbb{R}} h(x)\, k(y) \cdot p_X(x)\, p_Y(y) && \text{(by (ii))} \\
&= \left(\sum_{x\in\mathbb{R}} h(x)\, p_X(x)\right) \left(\sum_{y\in\mathbb{R}} k(y)\, p_Y(y)\right) \\
&= E(h(X)) \cdot E(k(Y)) && \text{(by (3.12))}
\end{aligned}$$

(iii) $\implies$ (i): Let $S, T \subseteq \mathbb{R}$. Consider the *indicator functions* $I_S : \mathbb{R} \to \{0, 1\}$ and $I_T : \mathbb{R} \to \{0, 1\}$ defined by:

$$I_S(x) \;=\; \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases} \qquad \text{and} \qquad I_T(x) \;=\; \begin{cases} 1 & \text{if } x \in T \\ 0 & \text{if } x \notin T \end{cases}$$

Then:

$$\begin{aligned}
\Pr(X \in S) \cdot \Pr(Y \in T) &= \left(\sum_{x \in S} p_X(x)\right)\left(\sum_{y \in T} p_Y(y)\right) \\
&= \left(\sum_{x \in \mathbb{R}} I_S(x) p_X(x)\right)\left(\sum_{y \in \mathbb{R}} I_T(y) p_Y(y)\right) \\
&= E(I_S(X)) \cdot E(I_T(Y)) \\
&= E(I_S(X) \cdot I_T(Y)) \qquad \text{(by (iii))} \\
&= \sum_{(x,y) \in \mathbb{R}^2} I_S(x) \cdot I_T(y) \cdot p_{XY}(x,y) \\
&= \sum_{x \in S} \sum_{y \in T} p_{XY}(x,y) \\
&= \Pr(X \in S, Y \in T)
\end{aligned}$$

□

(3.29) Corollary.

(i) Let X, Y be independent random variables. Then:

$$E(XY) = E(X) \cdot E(Y)$$

(ii) Let X, Y be independent random variables with moment generating functions $\phi_X(t)$ and $\phi_Y(t)$. Then:

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t)$$

(iii) Let $X_1, \ldots, X_n$ be independent random variables having all the same distribution. Then:

$$\phi_{X_1+\cdots+X_n}(t) = (\phi_{X_1}(t))^n$$

(3.30) Definition. The *correlation* between two random variables X and Y is defined to be $E(XY)$.

(3.31) Lemma (Cauchy–Schwarz inequality). Let X, Y be random variables. Then:

$$|E(XY)| \leq \sqrt{E(X^2) \cdot E(Y^2)}$$

If $\mathcal{A} = \mathcal{P}(\Omega)$, then equality holds if and only if there exists some $\lambda \in \mathbb{R}$ with:

$$X(\omega) = \lambda \cdot Y(\omega) \qquad \text{for all } \omega \in \Omega^* := \{\omega \in \Omega \mid \Pr(\{\omega\}) > 0\}$$

$\left(\text{In this case: } \lambda = \frac{E(XY)}{E(Y^2)}\right)$

Proof: Put $\lambda := \frac{E(XY)}{E(Y^2)}$. Then:

$$
\begin{array}{rrcl}
 & 0 & \leq & E((X-\lambda Y)^2) \;=\; E(X^2 - 2\lambda XY + \lambda^2 Y^2) \\
\Longrightarrow & 0 & \leq & E(X^2) \;-\; 2\lambda\, E(XY) \;+\; \lambda^2\, E(Y^2) \\
\Longrightarrow & 0 & \leq & E(X^2) \;-\; 2\dfrac{(E(XY))^2}{E(Y^2)} \;+\; \dfrac{(E(XY))^2}{(E(Y^2))^2}\, E(Y^2) \\
\Longrightarrow & 0 & \leq & E(X^2) \;-\; \dfrac{(E(XY))^2}{E(Y^2)} \\
\Longrightarrow & (E(XY))^2 & \leq & E(X^2)\, E(Y^2) \\
\Longrightarrow & |E(XY)| & \leq & \sqrt{E(X^2)\cdot E(Y^2)}
\end{array}
$$

This proves the stated inequality.

Furthermore, reversing the steps above, equality in the last inequality also implies equality in the first inequality, i.e.:

$$E((X-\lambda Y)^2) \;=\; 0 \qquad \Longleftrightarrow \qquad X(\omega) \;=\; \lambda \cdot Y(\omega) \quad \text{for all } \; \omega \in \Omega^*$$

□

(3.32) Definition. Let X, Y be random variables and $m_X = E(X)$, ${\sigma_X}^2 = \operatorname{Var}(X)$, $m_Y = E(Y)$, ${\sigma_Y}^2 = \operatorname{Var}(Y)$. The *covariance* between X and Y is defined by:

$$\operatorname{Cov}(X,Y) \;:=\; E\left((X-m_X)(Y-m_Y)\right) \;=\; E(XY) - m_X\, m_Y$$

X and Y are said to be *uncorrelated* if $\operatorname{Cov}(X,Y) = 0$.

(3.33) Remark. Let X, Y be random variables. If X and Y are independent, then X and Y are also uncorrelated.

Proof: Follows from (3.29)(i). □

(3.34) Definition. Let X be a random variable whose expectation $m = E(X)$ and standard deviation σ_X exist and are finite. Then, the *normalized random variable* defined by X is:

$$X^o \;=\; \frac{X-m}{\sigma_X}$$

Note: $E(X^o) = 0$ and $\sigma_{X^o} = 1$.

(3.35) Definition. The *correlation coefficient* of X and Y is defined to be the correlation between the normalized random variables defined by X and Y:

$$\rho_{XY} \;:=\; E\left(\left(\frac{X-m_X}{\sigma_X}\right)\left(\frac{Y-m_Y}{\sigma_Y}\right)\right) \;=\; \frac{\operatorname{Cov}(X,Y)}{\sigma_X\,\sigma_Y}$$

(3.36) Lemma. Let X, Y be random variables. Then:

$$|\rho_{XY}| \leq 1$$

If $\mathcal{A} = \mathcal{P}(\Omega)$, then equality holds if and only if $X|_{\Omega^*}$ and $Y|_{\Omega^*}$ are related by a linear function plus a constant, where $\Omega^* := \{\omega \in \Omega \mid \Pr(\{\omega\}) > 0\}$.

Proof: Applying (3.31) to the normalized random variables defined by X and Y yields:

$$|\rho_{XY}| \leq \sqrt{E\left(\left(\frac{X - m_X}{\sigma_X}\right)^2\right) E\left(\left(\frac{Y - m_Y}{\sigma_Y}\right)^2\right)} = 1$$

Furthermore, $|\rho_{XY}| = 1$, if and only if

$$\frac{X|_{\Omega^*} - m_X}{\sigma_X} = \lambda \cdot \frac{Y|_{\Omega^*} - m_Y}{\sigma_Y}$$

for $\lambda = \rho_{XY} \in \{1, -1\}$. (See the proof of (3.31) for the calculation of λ.) Hence

$$X|_{\Omega^*} = \rho_{XY} \cdot \frac{\sigma_X}{\sigma_Y} \cdot Y|_{\Omega^*} + \left(m_X - \rho_{XY} \cdot \frac{\sigma_X}{\sigma_Y} m_Y\right)$$

if and only if $\rho_{XY} \in \{1, -1\}$. □

(3.37) Theorem. Let $X_1, \ldots, X_n$ be random variables. Then:

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2 \cdot \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j)$$

Proof:

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^n X_i\right) &= E\left(\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2\right) \\
&= E\left(\left(\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)\right)^2\right) \qquad \text{(by (3.25))} \\
&= E\left(\left(\sum_{i=1}^n (X_i - E(X_i))\right)^2\right) \\
&= E\left(\sum_{i=1}^n \sum_{j=1}^n (X_i - E(X_i)) \cdot (X_j - E(X_j))\right) \\
&= E\left(\sum_{i=1}^n (X_i - E(X_i))^2 + 2 \cdot \sum_{1 \leq i < j \leq n} (X_i - E(X_i))(X_j - E(X_j))\right)
\end{aligned}$$

$$= \sum_{i=1}^{n} E\left((X_i - E(X_i))^2\right) \ + \ 2 \cdot \sum_{1 \le i < j \le n} E\left((X_i - E(X_i))(X_j - E(X_j))\right)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) \ + \ 2 \cdot \sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j)$$

□

(3.38) Corollary. Let $X_1, \ldots, X_n$ be random variables, such that X_i and X_j are uncorrelated if $i \neq j$. Then:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \ = \ \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Skewness and Kurtosis

(3.39) Definition. Let X be a random variable with $m = E(X)$ and $\sigma^2 = \mathrm{Var}(X) > 0$. The third and fourth moment of the normalized random variable defined by X are the *skewness* and *kurtois* of X:

$$\mathrm{skew}(X) \ := \ E\left(\left(\frac{X - m}{\sigma}\right)^3\right)$$

$$\mathrm{kurt}(X) \ := \ E\left(\left(\frac{X - m}{\sigma}\right)^4\right)$$

The skewness measures the lack of symmetry, while the kurtosis measures the fatness in the tails of the pdf of X.

(3.40) Lemma.

(i) $\mathrm{skew}(X) \ = \ \dfrac{E\left(X^3\right) - 3m\sigma^2 - m^3}{\sigma^3}$

(ii) $\mathrm{kurt}(X) \ = \ \dfrac{E\left(X^4\right) - 4mE\left(X^3\right) + 6m^2\sigma^2 + 3m^4}{\sigma^4}$

Universally valid inequalities

(3.41) Markov inequality. If X is a nonnegative random variable and $a > 0$, then:

$$\Pr(X \geq a) \quad \leq \quad \frac{E(X)}{a}$$

Proof:

$$\begin{aligned} E(X) \quad &= \quad \sum_{x \in \mathbb{R}} x\, p_X(x) \quad \geq \quad \sum_{x \geq a} x\, p_X(x) \\ &\geq \quad \sum_{x \geq a} a\, p_X(x) \quad = \quad a \sum_{x \geq a} p_X(x) \quad = \quad a \Pr(X \geq a) \end{aligned}$$

□

(3.42) Chebyshev inequality. If X is a random variable and $a > 0$, then:

$$\Pr\big(|X| \geq a\big) \quad \leq \quad \frac{E(X^2)}{a^2}$$

Furthermore, if $m = E(X)$ is finite, then:

$$\Pr\big(|X - m| \geq a\big) \quad \leq \quad \frac{\mathrm{Var}(X)}{a^2}$$

Proof:

$$\Pr\big(|X| \geq a\big) \quad = \quad \Pr(X^2 \geq a^2) \quad \overset{(3.41)}{\leq} \quad \frac{E(X^2)}{a^2}$$

$$\Pr\big(|X - m| \geq a\big) \quad \leq \quad \frac{E((X-m)^2)}{a^2} \quad = \quad \frac{\mathrm{Var}(X)}{a^2}$$

□

(3.43) Weak law of large numbers. Let $X_1, X_2, \ldots$ be a sequence of uncorrelated random variables with a common mean $m = E(X_i)$ and a common variance $\sigma^2 = \mathrm{Var}(X_i)$ for all $i \in \mathbb{N}$. For every $n \in \mathbb{N}$ the *sample mean* of the first n X_i's is defined by:

$$M_n \quad := \quad \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Then for any $\varepsilon > 0$:

$$\Pr\big(|M_n - m| \geq \varepsilon\big) \quad \leq \quad \frac{\sigma^2}{n\,\varepsilon^2} \longrightarrow 0 \qquad \text{for } n \to \infty$$

Proof:

$$\begin{aligned}
\Pr\big(|M_n - m| \geq \varepsilon\big) &= \Pr\big(|M_n - E(M_n)| \geq \varepsilon\big) \\
&\leq \frac{1}{\varepsilon^2}\,\mathrm{Var}(M_n) && \text{(by (3.42))} \\
&= \frac{1}{n^2\,\varepsilon^2}\,\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2\,\varepsilon^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) && \text{(by (3.38))} \\
&= \frac{\sigma^2}{n\,\varepsilon^2}
\end{aligned}$$

□

(3.44) Example. How often must a fair coin be flipped such that, with a probability at least 98%, “heads” show up in 49%–51% of all cases?

The task here is to find an n, such that:

$$\Pr\left(49\% \leq M_n \leq 51\%\right) \;\geq\; 98\%$$

$$\Longleftrightarrow \qquad \Pr\left(\left|M_n - \frac{1}{2}\right| \geq 1\%\right) \;\leq\; 2\%$$

From the weak law of large numbers (3.43) a number n is definitely big enough, if

$$\frac{\sigma^2}{n\,\varepsilon^2} \;\leq\; \frac{2}{100}\,,$$

where $\sigma^2 = \frac{1}{4}$ (see (3.48)(ii)) and $\varepsilon = 1\% = 0.01$. Hence, if a fair coin is flipped

$$n \;\geq\; 125{,}000$$

many times, “head” will show up in 49%–51% of all cases with a probability of at least 98%.

3.1 Uniform random variables

(3.45) Definition. The random variable X is a *uniform random variable*, uniformly distributed on N distinct numbers $\{x_1, \ldots, x_N\}$, if its pmf is given by:

$$p_X(x_i) \;:=\; \frac{1}{N} \qquad \text{for } \; i = 1, \ldots, N$$

(3.46) Lemma. If X is uniformly distributed random variable on N distinct numbers $\{x_1, \ldots, x_N\}$, then:

(i) $E(X) \;=\; \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} x_i$

(ii) $\mathrm{Var}(X) \;=\; \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} x_i^2 \;-\; \dfrac{1}{N^2} \left(\sum_{i=1}^{N} x_i \right)^2$

3.2 Bernoulli random variables

(3.47) Definition. A *Bernoulli random variable* with parameter $p \in (0,1)$ is a random variable X having a pmf given by:

$$p_X(0) \;:=\; 1 - p\,, \qquad p_X(1) \;:=\; p$$

This may be denoted by $X \sim \mathrm{Bernoulli}(p)$.

(3.48) Lemma. If $X \sim \mathrm{Bernoulli}(p)$, then:

(i) $E(X) \;=\; p$

(ii) $\mathrm{Var}(X) \;=\; p(1-p)$

(iii) $\phi_X(t) \;=\; (1-p) \,+\, p\, e^t \qquad$ for all $t \in \mathbb{R}$

Proof:

(i) $E(X) \;=\; 0 \cdot (1-p) \,+\, 1 \cdot p \;=\; p$

(ii) $\mathrm{Var}(X) \;=\; E((X - E(X))^2) \;=\; (-p)^2 \cdot (1-p) \,+\, (1-p)^2 \cdot p \;=\; p(1-p)$

(iii) $\phi_X(t) \;=\; e^{t \cdot 0} \cdot (1-p) \,+\, e^{t \cdot 1} \cdot p \;=\; (1-p) + p\, e^t$

$\square$

3.3 Binomial random variables

If $X_1, \dots, X_n$ are independent Bernoulli(p) random variables, then the sum

$$X_1 + \dots + X_n$$

is a binomial(n, p) random variable according to the following definition.

(3.49) Definition. A *binomial random variable* is a random variable X having a distribution given by:

$$p_X(i) \;:=\; \binom{n}{i} p^i (1-p)^{n-i} \qquad \text{for } \; i = 0, 1, \dots, n$$

where $n \in \mathbb{N}$ and $p \in (0, 1)$ are fixed parameters. This may be denoted by $X \sim \text{binomial}(n, p)$.

(3.50) Lemma. If $X \sim \text{binomial}(n, p)$, then:

(i) $E(X) \;=\; np$

(ii) $\text{Var}(X) \;=\; np(1-p)$

(iii) $\phi_X(t) \;=\; ((1-p) + pe^t)^n \quad$ for all $t \in \mathbb{R}$

Proof: Let

$$X \;=\; X_1 + \dots + X_n$$

with independent random variables $X_i \sim \text{Bernoulli}(p)$.

(i)

$$E(X) \;=\; E(X_1 + \dots + X_n) \;=\; E(X_1) + \dots + E(X_n) \;=\; np$$

(ii) As the X_i's are independent, it follows that:

$$\text{Var}(X) \;=\; \text{Var}(X_1 + \dots + X_n) \;=\; \text{Var}(X_1) + \dots + \text{Var}(X_n) \;=\; np(1-p)$$

(iii)

$$\phi_X(t) \;=\; E(e^{tX}) \;=\; E(e^{t(X_1 + \dots + X_n)}) \;=\; E(e^{tX_1} \cdot e^{tX_2} \dots e^{tX_n})$$

Furthermore, as the X_i's are independent, it follows that:

$$\phi_X(t) \;\;=\;\; E(e^{tX_1}) \cdot E(e^{tX_2}) \dots E(e^{tX_n}) \;\;=\;\; ((1-p) \,+\, pe^t)^n$$

□

The probability mass and cumulative distribution functions of a random variable X with $X \sim \text{binomial}(50, 0.5)$ are displayed in Figure 14.

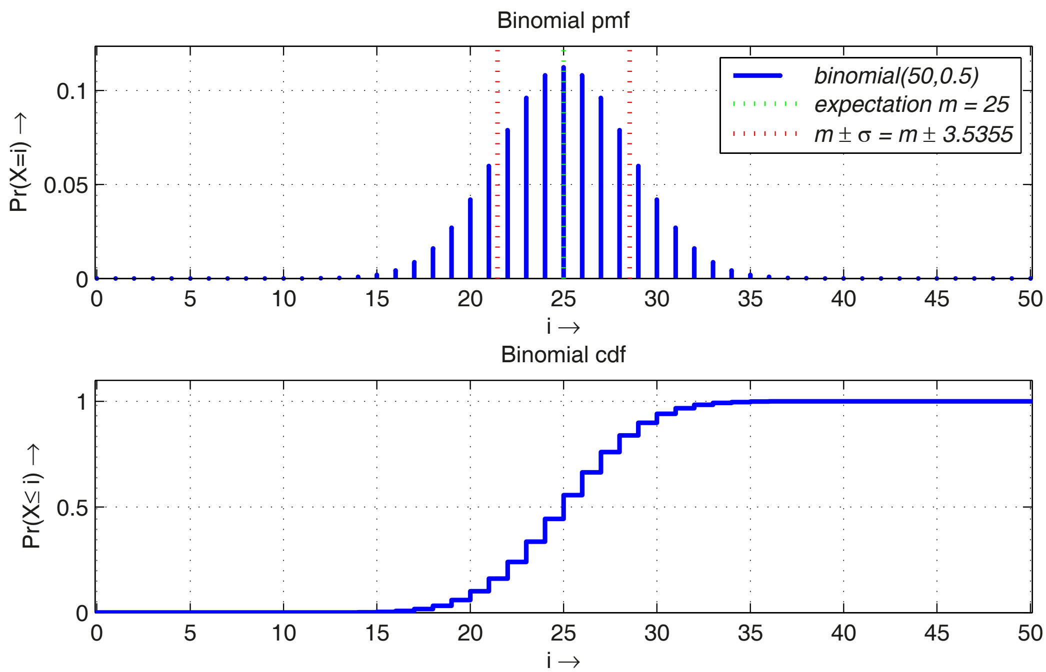

Figure 14: Binomial probability mass and cumulative distribution functions

3.4 Geometric random variables

(3.51) Definition. Let $p \in (0,1)$. A random variable X with a distribution given by

$$p_X(i) \;:=\; (1-p)^{i-1}p \qquad \text{for } \; i \in \mathbb{N}$$

is called a *geometric random variable*. This may be denoted by $X \sim \text{geometric}(p)$.

(3.52) Lemma. If $X \sim \text{geometric}(p)$, then:

(i) $E(X) \;=\; \dfrac{1}{p}$

(ii) $\mathrm{Var}(X) \;=\; \dfrac{1-p}{p^2} \;=\; \dfrac{1}{p^2} - \dfrac{1}{p}$

(iii) $\phi_X(t) \;=\; \dfrac{p\,e^t}{1-(1-p)\,e^t} \qquad \text{for } \; t < \ln\left(\dfrac{1}{1-p}\right)$

<u>*Proof:*</u>

(i),(ii): Taking the first and second derivative of the identity for the geometric series

$$\frac{1}{1-x} \;=\; \sum_{i=0}^{\infty} x^i$$

gives

$$\frac{1}{(1-x)^2} \;=\; \sum_{i=1}^{\infty} i x^{i-1}$$

and

$$\frac{2}{(1-x)^3} \;=\; \sum_{i=2}^{\infty} i(i-1)x^{i-2} \;=\; \sum_{i=1}^{\infty} (i+1)ix^{i-1} \;=\; \sum_{i=1}^{\infty} i^2 x^{i-1} + \frac{1}{(1-x)^2}$$

for $x < 1$. Using these identities, it follows that:

$$E(X) \;=\; \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1}p \;=\; p \cdot \sum_{i=1}^{\infty} i \cdot (1-p)^{i-1} \;=\; p \cdot \frac{1}{(1-(1-p))^2} \;=\; \frac{1}{p}$$

and

$$E(X^2) \;=\; \sum_{i=1}^{\infty} i^2 \cdot (1-p)^{i-1}p \;=\; p\left(\frac{2}{(1-(1-p))^3} - \frac{1}{(1-(1-p))^2}\right) \;=\; \frac{2-p}{p^2}$$

$$\Longrightarrow \qquad \mathrm{Var}(X) \;=\; E(X^2) - (E(X))^2 \;=\; \frac{2-p}{p^2} - \frac{1}{p^2} \;=\; \frac{1-p}{p^2} \;=\; \frac{1}{p^2} - \frac{1}{p}$$

(iii):

$$\phi_X(t) \;=\; \sum_{i=1}^{\infty} e^{t\cdot i} \cdot (1-p)^{i-1} \cdot p \;=\; pe^t \sum_{i=1}^{\infty} e^{t\cdot(i-1)} \cdot (1-p)^{i-1}$$

$$\;=\; pe^t \sum_{i=0}^{\infty} ((1-p)e^t)^i \;=\; \frac{p\,e^t}{1-(1-p)\,e^t} \qquad \square$$

(3.53) Exercise. Let $p \in (0, 1)$ and X a random variable with $X \sim \text{geometric}(p)$. Use the moment generating function $\phi_X(t)$ given in (3.52)(iii) to provide alternative proofs of (3.52)(i) and (3.52)(ii) on the basis of (3.18).

The probability mass and cumulative distribution functions of a random variable X with $X \sim \text{geometric}(0.1)$ are displayed in Figure 15.

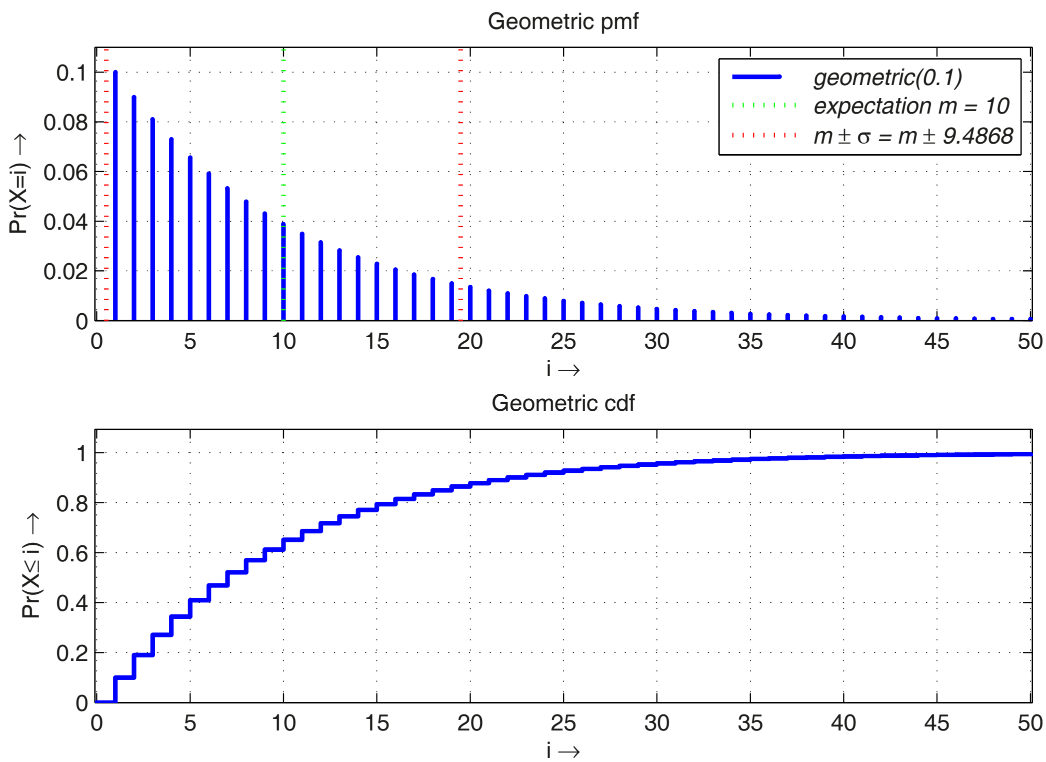

Figure 15: Geometric probability mass and cumulative distribution functions

3.5 Negative binomial random variables

(3.54) Definition. Let $p \in (0,1)$ and $n \in \mathbb{N}$. A *negative binomial random variable* (*Pascal random variable*) with parameters p and n is a random variable X having a distribution given by:

$$p_X(i) := \binom{i-1}{n-1} p^n (1-p)^{i-n} \qquad \text{for all } \; i \in \{n, n+1, \dots\}$$

This may be denoted by $X \sim \text{nbino}(n,p)$.

(3.55) Lemma. If $X \sim \text{nbino}(n,p)$, then:

(i) $E(X) \;=\; \dfrac{n}{p}$

(ii) $\text{Var}(X) \;=\; n\,\dfrac{1-p}{p^2} \;=\; \dfrac{n}{p^2} - \dfrac{n}{p}$

(iii) $\phi_X(t) \;=\; \left(\dfrac{p\,e^t}{1-(1-p)\,e^t}\right)^n \qquad \text{for } \; t < \ln\left(\dfrac{1}{1-p}\right)$

The probability mass functions for some negative binomial distributions are displayed in Figure 16.

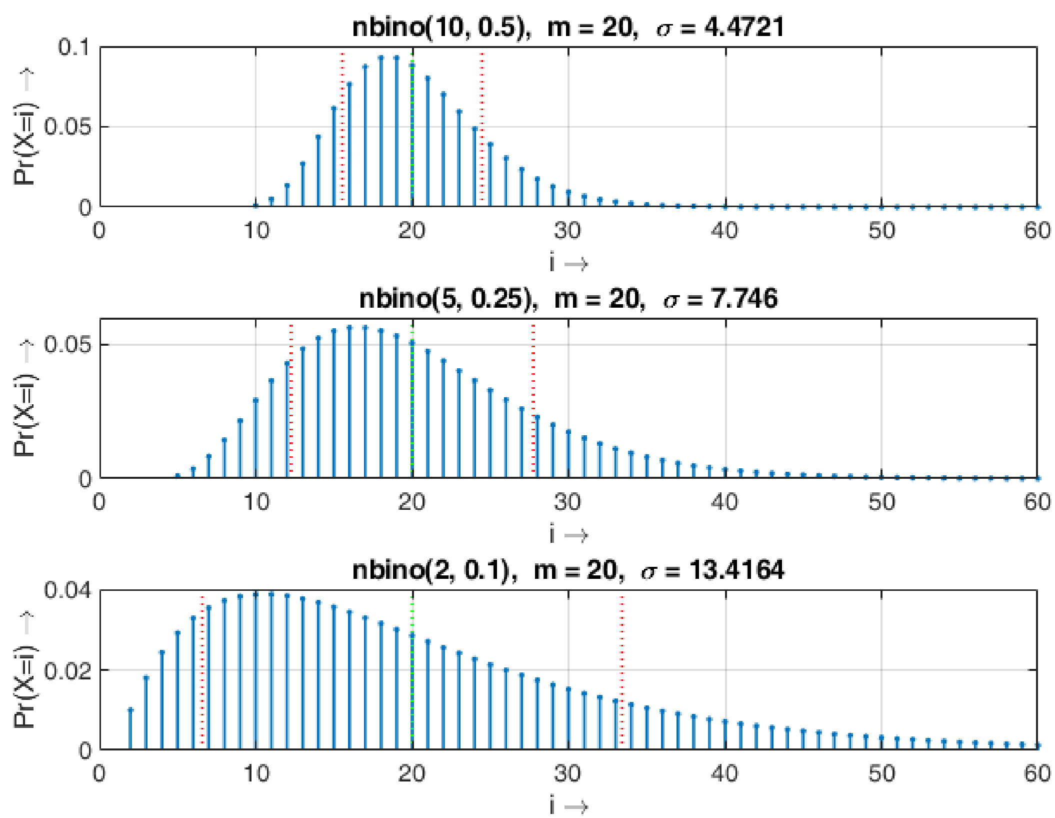

Figure 16: Negative binomial probability mass functions with $np = 20$

3.6 Poisson random variables

(3.56) Definition. A *Poisson random variable* is a random variable X having a distribution given by

$$p_X(i) \; := \; e^{-\lambda} \frac{\lambda^i}{i!} \qquad \text{for all} \;\; i \in \mathbb{N}_0,$$

where $\lambda > 0$ is some fixed parameter. This may be denoted by $X \sim \text{Poisson}(\lambda)$.

(3.57) Lemma. If $X \sim \text{Poisson}(\lambda)$, then:

(i) $E(X) \; = \; \lambda$

(ii) $\text{Var}(X) \; = \; \lambda$

(iii) $\phi_X(t) \; = \; e^{\lambda(e^t-1)} \quad$ for all $t \in \mathbb{R}$

<u>*Proof:*</u>

(i)

$$E(X) \; = \; \sum_{i=0}^{\infty} i e^{-\lambda} \frac{\lambda^i}{i!} \; = \; \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \; = \; \lambda e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \; = \; \lambda \, e^{-\lambda} \, e^{\lambda} \; = \; \lambda$$

(ii)

$$\begin{aligned}
\text{Var}(X) \; &= \; E(X^2) \; - \; (E(X))^2 \\
&= \; \sum_{i=0}^{\infty} i^2 e^{-\lambda} \frac{\lambda^i}{i!} \; - \; \lambda^2 \\
&= \; \lambda \, e^{-\lambda} \sum_{i=1}^{\infty} i \frac{\lambda^{i-1}}{(i-1)!} \; - \; \lambda^2 \\
&= \; \lambda \, e^{-\lambda} \left(\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} + \sum_{i=1}^{\infty} (i-1) \, \frac{\lambda^{i-1}}{(i-1)!} \right) \; - \; \lambda^2 \\
&= \; \lambda \, e^{-\lambda} \left(e^{\lambda} + \lambda e^{\lambda} \right) \; - \; \lambda^2 \; = \; \lambda
\end{aligned}$$

(iii)

$$\phi_X(t) \; = \; \sum_{i=0}^{\infty} e^{t \cdot i} \, e^{-\lambda} \, \frac{\lambda^i}{i!} \; = \; e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} \; = \; e^{-\lambda} \, e^{\lambda e^t} \; = \; e^{\lambda(e^t-1)}$$

□

The probability mass and cumulative distribution functions of a random variable X with $X \sim \text{Poisson}(10)$ are displayed in Figure 17.

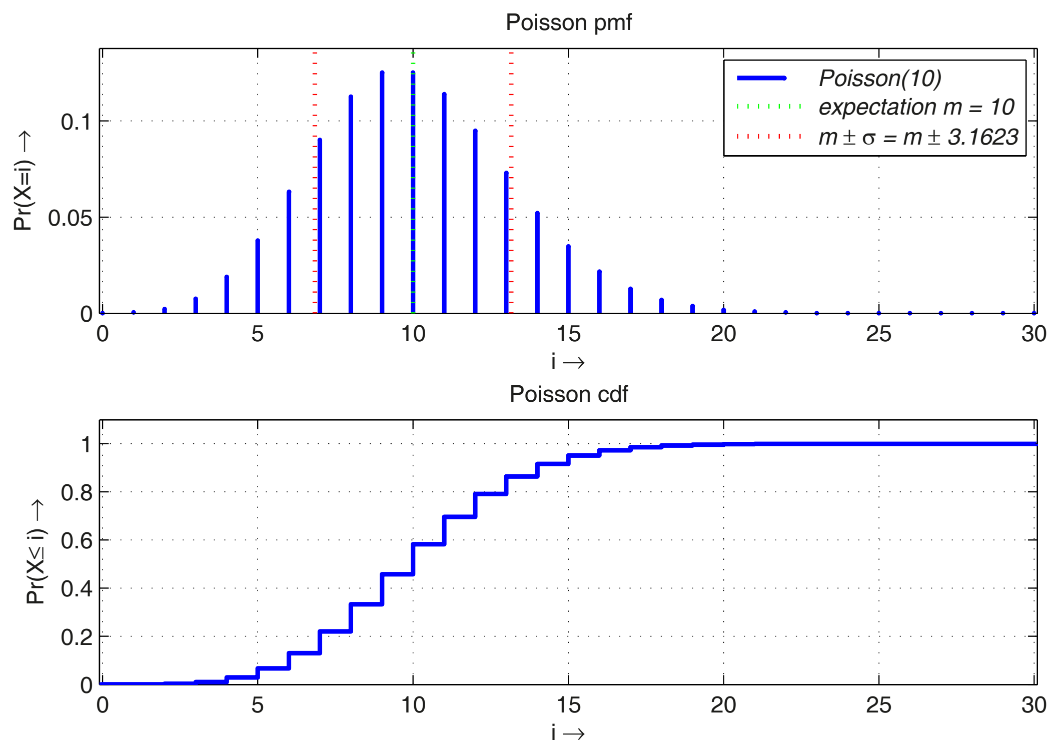

Figure 17: Poisson probability mass and cumulative distribution functions

(3.58) Poisson approximation of binomial probabilities. For small values of p ($p \lesssim 0.01$), the binomial distribution can be approximated by the Poisson distribution with mean $\lambda = np$. To be accurate:

(3.59) Theorem. If $(p_n)_{n\in\mathbb{N}}$ is a sequence of numbers with $p_n \in (0,1)$ and

$$\lim_{n\to\infty} n \cdot p_n \quad = \quad \lambda$$

for some $\lambda \in \mathbb{R}$, then the sequence of binomial distributions binomial(n, p_n) converges to the Poisson distribution Poisson(λ).

<u>*Proof:*</u>

$$\begin{aligned}
p_{\text{binomial}(n,p_n)}(i) &= \binom{n}{i} \cdot p_n^i (1-p_n)^{n-i} \\
&= \frac{n(n-1)\cdots(n-(i-1))}{i!} \cdot p_n^i (1-p_n)^{n-i} \\
&= \frac{(np_n)^i}{i!} \cdot \frac{n(n-1)\cdots(n-(i-1))}{n^i} \cdot (1-p_n)^{n-i} \\
&= \frac{(np_n)^i}{i!} \cdot \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{i-1}{n}\right) \cdot (1-p_n)^{-i} \left(1-\frac{np_n}{n}\right)^n \\
&\to \frac{\lambda^i}{i!} \cdot e^{-\lambda} \qquad \text{for } n\to\infty
\end{aligned}$$

□

Far better is the following result, which gives a quantitative error estimate for the cumulated differences between a binomoial pmf and its approximating Poisson pmf.

(3.60) Theorem. Let $n \in \mathbb{N}$ and $p \in (0,1)$. Then:

$$\sum_{i=0}^{\infty} \left| p_{\text{binomial}(n,p)}(i) - p_{\text{Poisson}(n \cdot p)}(i) \right| \leq 2np^2$$

(3.61) MatLab–App. In order to get an idea of the goodness of the Poisson approximation for binomial(n,p) a MatLab–App is provided (see Figure 18). (Download the files `BinominalVsPoisson.m` and `BinominalVsPoisson.fig` into your current MatLab folder and run `BinominalVsPoisson.m`.)

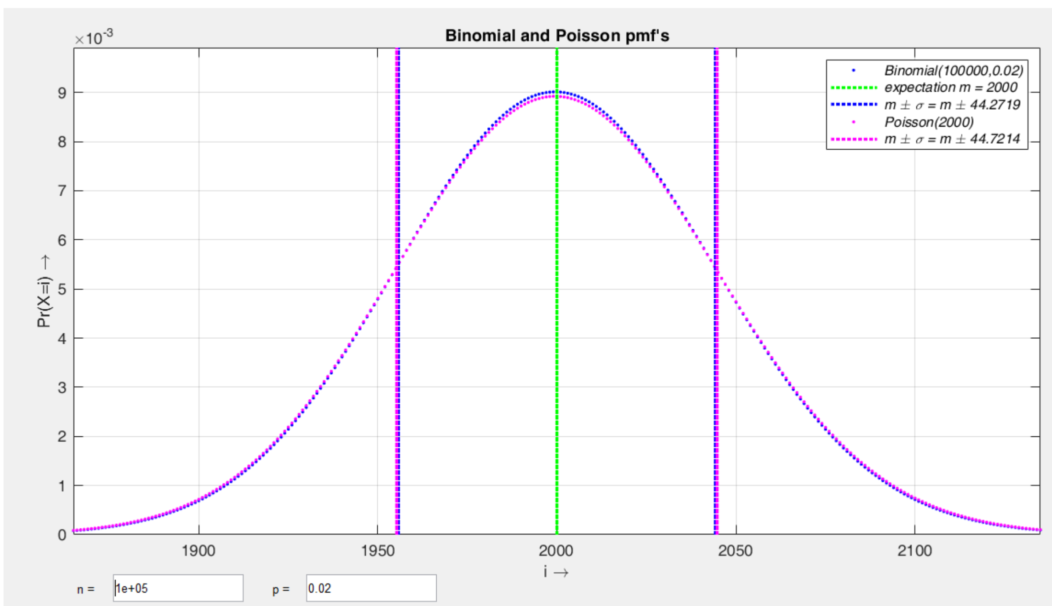

Figure 18: binomial(n,p) and Poisson(λ) with $\lambda = np$ for values in $[\lambda - 3\sqrt{\lambda}, \lambda + 3\sqrt{\lambda}]$

3.7 Hypergeometric random variables

Assume, a total number of N entities are given containing D defective elements. Furthermore assume that $n \leq N$ elements are drawn randomly without replacement. Let $X_i = 1$ if the i'th element drawn is defective and $X_i = 0$ otherwise. Then

$$X \;=\; X_1 + \cdots + X_n$$

is a hypergeometric random variable:

(3.62) Definition. A *hypergeometric random variable* is a random variable X having a distribution given by

$$p_X(i) \;:=\; \frac{\binom{D}{i}\binom{N-D}{n-i}}{\binom{N}{n}} \qquad \text{for } i = 0, 1, \ldots, n,$$

where $N, D, n \in \mathbb{N}$ are fixed parameters with $D \leq N$ and $n \leq N$. This may be denoted by $X \sim \text{hypergeometric}(N, D, n)$.

(3.63) Lemma. If $X \sim \text{hypergeometric}(N, D, n)$ and $p := \frac{D}{N}$, then:

(i) $E(X) \;=\; \dfrac{n\,D}{N} \;=\; np$

(ii) $\text{Var}(X) \;=\; \dfrac{nD(N-D)}{N^2}\left(1 - \dfrac{n-1}{N-1}\right) \;=\; np(1-p) \cdot \left(1 - \dfrac{n-1}{N-1}\right)$

<u>*Proof:*</u>

(i)

$$E(X) \;=\; E(X_1 + \cdots + X_n) \;=\; E(X_1) + \cdots + E(X_n) \;=\; n\,\frac{D}{N}$$

(ii) (3.37) gives

$$\begin{aligned} \text{Var}(X) \;&=\; \sum_{i=1}^{n} \text{Var}(X_i) \;+\; 2 \cdot \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \\ &=\; n \cdot \frac{D}{N} \cdot \frac{N-D}{N} \;+\; 2 \cdot \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \end{aligned}$$

For $i \neq j$, $X_i \cdot X_j$ is a Bernoulli variable with:

$$\Pr(X_i \cdot X_j = 1) \;=\; \frac{D(D-1)}{N(N-1)}$$

Therefore

$$\begin{aligned} \text{Cov}(X_i, X_j) \;&=\; E(X_i X_j) - E(X_i)E(X_j) \\ &=\; \frac{D(D-1)}{N(N-1)} - \left(\frac{D}{N}\right)^2 \;=\; \frac{-D(N-D)}{N^2(N-1)} \end{aligned}$$

and it follows that:

$$\text{Var}(X) \;=\; \frac{n\,D\,(N-D)}{N^2} \;-\; 2\binom{n}{2}\frac{D(N-D)}{N^2(N-1)} \;=\; \frac{n\,D\,(N-D)}{N^2}\left(1 - \frac{n-1}{N-1}\right)$$

□

$X \sim$	pmf	$E(X)$	$\mathrm{Var}(X)$	$\phi_X(t)$	$X_1 + \cdots + X_m \quad (X_i \sim X)$
Bernoulli(p)	$p_X(0) = 1-p\,, \quad p_X(1) = p$	p	$p(1-p)$	$(1-p) + p\,e^t$	binomial(m, p)
binomial(n, p)	$p_X(i) = \binom{n}{i}\, p^i (1-p)^{n-i} \qquad (i = 0, 1, \ldots, n)$ `binopdf(i,n,p)`	np	$np(1-p)$	$((1-p) + p\,e^t)^n$	binomial(mn, p)
geometric(p)	$p_X(i) = (1-p)^{i-1} p \qquad (i \in \mathbb{N})$ `geopdf(i-1,p)`	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{p\,e^t}{1-(1-p)\,e^t}$	nbino(m, p)
nbino(n, p)	$p_X(i) = \binom{i-1}{n-1} p^n (1-p)^{i-n} \qquad (i \geq n)$ `nbinpdf(i-n,n,p)`	$\frac{n}{p}$	$n \cdot \frac{1-p}{p^2}$	$\left(\frac{p\,e^t}{1-(1-p)\,e^t}\right)^n$	nbino(mn, p)
Poisson(λ) $\quad (\lambda > 0)$	$p_X(i) = e^{-\lambda}\, \frac{\lambda^i}{i!} \qquad (i \in \mathbb{N}_0)$ `poisspdf(i,lambda)`	λ	λ	$e^{\lambda(e^t - 1)}$	Poisson($m\lambda$)
hypergeometric(N, D, n) $(D, n \leq N)$	$p_X(i) = \frac{\binom{D}{i}\binom{N-D}{n-i}}{\binom{N}{n}} \qquad (i = 0, 1, \ldots, n)$ `hygepdf(i, N, D, n)`	$n\,\frac{D}{N}$	$\frac{nD(N-D)}{N^2}\left(1 - \frac{n-1}{N-1}\right)$		