Appendix A
Prerequisites

A.1 Generating functions

Since generating functions of distributions on $\overline{\mathbb{N}}_0 = \mathbb{N}_0 \cup \{\infty\}$ play a central role in this text, we collect some of their most important properties in this appendix. We begin with the definition for the case of proper distributions on $\mathbb{N}_0$.

Definition A.1. Given a random variable X taking values in $\mathbb{N}_0$ and with distribution $Q = (p_n)_{n\geq 0}$, the *generating function (gf) of X or Q* is defined as

$$f(s) \ := \ \mathbb{E}s^X \ = \ \sum_{n\geq 0} p_n s^n \quad \text{for } |s| \leq R_Q,$$

where $R_Q \geq 1$ denotes the radius of convergence of the right-hand power series.

Although formulated for distributions on $\mathbb{N}_0$, which are our main concern here, the same definition holds for the gf of any finite, or even σ-finite measure Q on this set. Since $f(1) = \sum_{n\geq 0} p_n = Q(\mathbb{N}_0)$, we see that $R_Q \geq 1$ if $Q(\mathbb{N}_0)$ is finite, whereas even $R_Q = 0$ may hold if $Q(\mathbb{N}_0) = \infty$.

If $\varphi(t) = \mathbb{E}e^{-tX}$ for $t > \sigma_Q := \inf\{s \leq 0 : \mathbb{E}e^{-sX} < \infty\}$ denotes the Laplace transform (LT) of X, then the gf f can be directly derived from it by a logarithmic transform of its argument, viz. $f(s) = \varphi(-\log s)$ for $|s| < R_Q = e^{-\sigma_Q}$. Furthermore, f is actually defined on the whole complex disc $\mathbb{D}_{R_Q} = \{z \in \mathbb{C} : |z| \leq R_Q\}$, where it constitutes a continuous function [☞ Prop. A.4]. On the interior of $\mathbb{D}_{R_Q}$ it is even analytic (holomorphic). On the other hand, our analysis will mostly deal with f only on the subdomain $[0,1]$ in which case the definition of gf's may easily be extended to random variables on $\overline{\mathbb{N}}_0$ under the usual convention that $s^\infty := 0$ for $s \in [0,1)$ and $s^\infty := 1$ if $s = 1$.

Definition A.2. Given a random variable X taking values in $\overline{\mathbb{N}}_0$ and with distribution $Q = (p_n)_{n\in\overline{\mathbb{N}}_0}$, the *generating function (gf) of X or Q* is defined as

$$f(s) \;:=\; \mathbb{E}s^X \;=\; \sum_{n\geq 0} p_n s^n \quad \text{for } s\in[0,1].$$

Notice that f, when restricted to $[0,1)$, coincides with the gf f^*, say, of the defective distribution $Q^* = (p_n)_{n\geq 0}$ on $\mathbb{N}_0$. However, we have $f^*(1) = \sum_{n\geq 0} p_n = Q^*(\mathbb{N}_0)$, while $f(1) = 1 > f^*(1) = f(1-)$ whenever $p_\infty > 0$. This shows that f^* is continuous at 1 while f is so only if $p_\infty = 0$.

It is a well-known fact in complex analysis that an analytic function is uniquely determined by its values on a set $\{s_n : n \geq 0\}$ with accumulation point in the interior of its domain. Therefore, the following uniqueness theorem is immediate.

Theorem A.3. [Uniqueness theorem] *Let f_1, f_2 be the gf's of distributions Q_1, Q_2 on $\overline{\mathbb{N}}_0$. Then the following assertions are equivalent:*

(a) $f_1 = f_2$.
(b) $f_1(s) = f_2(s)$ *for all $s\in I$, where I is a subset of* $[0,1)$ *with accumulation point in* $[0,1)$.
(c) $Q_1 = Q_2$.

The most important analytic properties of gf's are listed in the following proposition.

Proposition A.4. *Let X be a random variable taking values in $\overline{\mathbb{N}}_0$ and with distribution $Q = (p_n)_{n\in\overline{\mathbb{N}}_0}$, The the following assertions hold true for its gf $f(s) = \mathbb{E}s^X$:*

(a) *For any $R < R_Q$, f is uniformly continuous on $\mathbb{D}_R$ with $|f(z)| \leq f(R)$.*
(b) *For any $R \leq R_Q$, f is analytic (holomorphic) on the interior of $\mathbb{D}_R$ with n^{th} derivative*

$$f^{(n)}(z) \;=\; n!\sum_{k\geq n}\binom{k}{n} p_k z^{k-n} \;=\; n!\,\mathbb{E}\left[\binom{X}{n} z^{X-n}\right].$$

(c) $f^{(n)}(1-) = \lim_{s\uparrow 1} f^{(n)}(s)$ *exists for all $n\in\mathbb{N}_0$ and is given by the n^{th} factorial moment of X, viz.*

$$f^{(n)}(1-) \;=\; \mathbb{E}X(X-1)\cdot\ldots\cdot(X-n+1) \;=\; n!\,\mathbb{E}\binom{X}{n},$$

where $f^{(0)} := f$ *and* $\binom{k}{n} := 0$ *if* $k < n$. *In particular,* $\mathbb{E}X = f'(1-)$, $\mathbb{E}X^2 = f''(1-) + f'(1-)$ *and* $\mathbb{V}\mathrm{ar}X = f''(1-) + f'(1-)(1-f'(1-))$. *If* $p_\infty = \mathbb{P}(X=\infty) = 0$, *then* $f^{(n)}(1-) = f^{(n)}(1)$ *for all* $n \in \mathbb{N}_0$.

(d) *For each* $n \in \mathbb{N}_0$, $f^{(n)}$ *is nondecreasing and convex on* $[0,R_Q)$, *and it is increasing if* $\mathbb{P}(X>n)>0$, *and strictly convex if* $\mathbb{P}(X>n+1)>0$.

Finally, we want to present a continuity theorem for gf's that provides the connection between the vague ($\stackrel{v}{\to}$) or weak ($\stackrel{w}{\to}$) convergence of distributions on $\mathbb{N}_0$ and the pointwise convergence of their gf's.

Theorem A.5. *Let* $Q_1, Q_2, \ldots$ *be distributions on* $\mathbb{N}_0$ *with gf's* $f_1, f_2, \ldots$

(a) *If* $Q_n \stackrel{v}{\to} Q$ *for a possibly defective distribution* Q *with gf* f, *then* $\lim_{n\to\infty} f_n(s) = f(s)$ *for all* $s \in [0,1)$.

(b) *If* $Q_n \stackrel{w}{\to} Q$ *for a distribution* Q *with gf* f, *then* $\lim_{n\to\infty} f_n(s) = f(s)$ *for all* $s \in [0,1]$, *and the convergence is uniform.*

(c) *Conversely, if* $f(s) := \lim_{n\to\infty} f_n(s)$ *exists for all* $s \in [0,1]$, *then* f *is the gf of a possibly defective distribution* Q *and* $Q_n \stackrel{v}{\to} Q$. *If* f *continuous at* 1, *i.e.* $f(1-) = f(1)$, *then* $Q(\mathbb{N}_0) = 1$ *and the convergence holds true in the weak sense, i.e.* $Q_n \stackrel{w}{\to} Q$.

A.2 Total variation distance and coupling

Let $\mathscr{M}_\pm = \mathscr{M}_\pm(\mathfrak{X},\mathscr{A})$ denote the space of finite signed measures on a given a measurable space $(\mathfrak{X},\mathscr{A})$, i.e. the vector space of all differences $\lambda - \mu$ of finite measures λ, μ on this space. Endowed with the supremum norm $\|\cdot\|$, that is

$$\|\lambda\| \;:=\; \sup_{A\in\mathscr{A}} |\lambda(A)|,$$

$(\mathscr{M}_\pm, \|\cdot\|)$ becomes a complete normed space (Banach space). The induced metric d_{tv} is called *total variation distance*. Clearly, *convergence in total variation* of λ_n to λ means uniform convergence, that is

$$\lim_{n\to\infty} \sup_{A\in\mathscr{A}} |\lambda_n(A) - \lambda(A)| \;=\; 0.$$

By a standard extension argument the latter condition is further equivalent to

$$\lim_{n\to\infty} \sup_{g\in b\mathscr{A}:\|g\|_\infty\le 1} \left| \int g\, d\lambda_n - \int g\, d\lambda \right|, \tag{A.1}$$

where $b\mathscr{A}$ denotes the space of bounded $\mathscr{A}$-measurable functions $g : \mathfrak{X} \to \mathbb{R}$.

Confining to probability distributions λ, μ, we have

$$\|\lambda - \mu\| \;=\; \int_{\{f>g\}} (f-g)\, d\nu \;=\; \int_{\{f<g\}} (g-f)\, d\nu \;=\; \frac{1}{2}\int |f-g|\, d\nu, \tag{A.2}$$

where f, g are the densities of λ, μ with respect to an arbitrary dominating measure ν (e.g. $\lambda + \mu$). With the help of this fact, we immediately get the following result for the set $\mathscr{P}(\mathfrak{X}, \mathscr{A})$ of probability measures on $(\mathfrak{X}, \mathscr{A})$.

Theorem A.6. *The metric space $(\mathscr{P}(\mathfrak{X}, \mathscr{A}), d_{tv})$ is complete.*

Proof. Given a Cauchy sequence $(\lambda_n)_{n\ge1}$ in $(\mathscr{P}(\mathfrak{X}, \mathscr{A}), d_{tv})$, let μ be a dominating probability measure, e.g. $\mu = \sum_{n\ge1} 2^{-n}\lambda_n$, and f_n a μ-density of λ_n for any $n \ge 1$. By (A.2),

$$\|\lambda_n - \lambda_m\| \;=\; \frac{1}{2}\|f_n - f_m\|_1 \;\to\; 0$$

as $m, n \to \infty$, where $\|\cdot\|_1$ is the usual L^1-norm on the space $L^1(\mu)$ of μ-integrable functions (modulo μ-a.s. equality). Since this space is a Banach space, we infer the existence of some (nonnegative) $f \in L^1(\mu)$ such that $\|f_n - f\|_1 \to 0$ and thus $\|f\|_1 = 1$. Putting $\lambda := f\mu$, we finally obtain

$$\|\lambda_n - \lambda\| \;=\; \frac{1}{2}\|f_n - f\|_1 \;\to\; 0$$

as $n \to \infty$ and therefore $d_{tv}(\lambda_n, \lambda) \to 0$. □

In the discrete case, where λ, μ are distributions on $\mathbb{N}_0$ (or any countable $\mathfrak{X}$), (A.2) becomes

$$\|\lambda - \mu\| \;=\; \frac{1}{2}\sum_{x\in\mathfrak{X}} |\lambda_x - \mu_x| \tag{A.3}$$

when choosing f, g as counting densities and setting $\lambda_x = \lambda(\{x\})$. Note further that λ, μ are mutually singular if, and only if, $\|\lambda - \mu\| = 1$ or, equivalently, $\lambda \wedge \mu = 0$, where

$$\lambda \wedge \mu(dx) \;:=\; (f \wedge g)(x)\, \nu(dx).$$

In general, we have by (A.2) that

$$\|\lambda - \mu\| \;=\; \int_{\{f>g\}} (f - f\wedge g)\, d\nu \;=\; \int (f - f\wedge g)\, d\nu \;=\; 1 - \|\lambda \wedge \mu\|. \tag{A.4}$$

A pair (X, Y) of random variables, defined on a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, is called a *coupling for* λ *and* μ if $X \stackrel{d}{=} \lambda$ and $Y \stackrel{d}{=} \mu$. It satisfies the *coupling inequality*

$$\|\lambda - \mu\| \;\le\; \mathbb{P}(X \neq Y), \tag{A.5}$$

because, for all $A \in \mathscr{A}$,

$$\begin{aligned}|\lambda(A) - \mu(A)| \;&=\; |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\ &=\; |\mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y)| \;\le\; \mathbb{P}(X \neq Y).\end{aligned}$$

The following lemma shows that there is always a coupling that provides equality in (A.5).

Lemma A.7. (Maximal coupling lemma)*Given two distributions λ, μ on a measurable space $(\mathfrak{X}, \mathscr{A})$ there exist random variables X, Y on a common probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ such that $X \stackrel{d}{=} \lambda$, $Y \stackrel{d}{=} \mu$ and*

$$\|\lambda - \mu\| \;=\; \mathbb{P}(X \neq Y).$$

*The pair (X, Y) is called a **maximal coupling of λ and μ**.*

Proof. Put $\alpha := \|\lambda \wedge \mu\|$, $\varphi := \alpha^{-1}(\lambda \wedge \mu) \in \mathscr{P}(\mathfrak{X})$ and observe that

$$\lambda \;=\; \alpha\varphi + (1-\alpha)\lambda' \quad \text{and} \quad \mu \;=\; \alpha\varphi + (1-\alpha)\mu'$$

with obviously defined $\lambda', \mu' \in \mathscr{P}(\mathfrak{X})$. Now let η, X', Y' and Z be independent random variables on a common probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ such that $X' \stackrel{d}{=} \lambda'$, $Y' \stackrel{d}{=} \mu'$, $Z \stackrel{d}{=} \varphi$ and $\mathbb{P}(\eta = 1) = 1 - \mathbb{P}(\eta = 0) = \alpha$. Defining

$$X \;:=\; \begin{cases} Z, & \text{if } \eta = 1, \\ X', & \text{if } \eta = 0 \end{cases} \quad \text{and} \quad Y \;:=\; \begin{cases} Z, & \text{if } \eta = 1, \\ Y', & \text{if } \eta = 0 \end{cases}$$

it is easily seen that $X \stackrel{d}{=} \lambda$, $Y \stackrel{d}{=} \mu$, and

$$\mathbb{P}(X \neq Y) \;=\; \mathbb{P}(\eta = 0) \;=\; 1 - \alpha \;=\; 1 - \|\lambda \wedge \mu\|$$

which, by (A.4), proves the assertion. □

Let us return to the special case of distributions on $\mathbb{N}_0$ and finally point out that for those weak convergence and total variation convergence (i.e. uniform convergence) are actually equivalent.

Theorem A.8. *Let $(\lambda_n)_{n\ge 0}$ be a sequence of distributions on $\mathbb{N}_0$. Then $\lambda_n \stackrel{w}{\to} \lambda_0$ holds iff $\|\lambda_n - \lambda_0\| \to 0$ as $n \to \infty$.*

Proof. We must only verify that $\lambda_n \xrightarrow{w} \lambda_0$ implies $\|\lambda_n - \lambda_0\| \to 0$. Hence assume weak convergence, fix any $\varepsilon \in (0,1)$ and choose N so large that

$$\sup_{n\geq 0} \lambda_n([N+1,\infty)) \;=\; \sup_{n\geq 0} \sum_{k>N} \lambda_{n,k} \;<\; \varepsilon,$$

where $\lambda_{n,k} = \lambda_n(\{k\})$. This is possible by tightness of the λ_n. Using (A.3), we now infer that

$$\|\lambda_n - \lambda_0\| \;\leq\; \frac{\varepsilon}{2} \;+\; \frac{1}{2}\sum_{k=0}^{N} |\lambda_{n,k} - \lambda_{0,k}|.$$

But the last sum converges to 0, for weak convergence on $\mathbb{N}_0$ clearly entails pointwise convergence of the counting densities. Hence $\|\lambda_n - \lambda_0\| \to 0$, for $\varepsilon \in (0,1)$ was chosen arbitrarily. □