Appendix

A1 Standard Distributions

We give here a list of standard distributions, for the main purpose of introducing the terminology and parameterization used in the book. For a discrete distribution on $\mathbb{N}$, $p(k)$ denotes the probability mass function and $\widehat{p}[z] \stackrel{\text{def}}{=} \sum_0^\infty z^k p(k)$ the probability generating function (p.g.f.). In the continuous case, $f(x)$ is the density, $F(x) \stackrel{\text{def}}{=} \int_{-\infty}^x f(y)\,\mathrm{d}y$ the c.d.f., $\widehat{F}[s] \stackrel{\text{def}}{=} \int_{-\infty}^\infty \mathrm{e}^{sy} f(y)\,\mathrm{d}y$ the m.g.f. (defined for any $s \in \mathbb{C}$ such that $\int_{-\infty}^\infty \mathrm{e}^{\Re s\, y} f(y)\,\mathrm{d}y < \infty$), and the cumulant function (c.g.f.) is $\log \widehat{F}[s]$.

The Bernoulli(p) **distribution** is the distribution of $X \in \{0,1\}$, such that $\mathbb{P}(X = 1) = 1$. Here the event $\{X = 1\}$ can be thought of as heads coming up when throwing a coin w.p. 1 for heads. The **binomial** (n,p) **distribution** is the distribution of the sum $N \stackrel{\text{def}}{=} X_1 + \cdots + X_n$ of n i.i.d. Bernoulli(p) r.v.'s. The p.g.f.'s are $\mathbb{E}z^X = 1 - p + pz$, $\mathbb{E}z^N = (1 - p + pz)^n$.

The geometric distribution with success parameter $1-\rho$ is the distribution of the number N of tails before a head comes up when flipping a coin w.p. ρ for tails, $\mathbb{P}(N = n) = (1-\rho)\rho^n$, $n \in \mathbb{N}$. Also $N' = N+1$, the total number of flips including the final head, is said to have a geometric($1-\rho$) distribution, and one has $\mathbb{P}(N' = n) = (1-\rho)\rho^{n-1}$, $n = 1, 2, \ldots$.

The Gamma(α, λ) **distribution** has

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

The m.g.f. is $\bigl(\lambda/(\lambda - s)\bigr)^\alpha$, the mean is α/λ, and the variance is α/λ^2. The Gamma$(1, \lambda)$ distribution is the exponential distribution with rate parameter λ, and the Gamma(n, λ) distribution with $n = 1, 2, 3, \ldots$ is commonly called the Erlang(n, λ) distribution or just the Erlang(n) distribution.

The **inverse Gamma**(α, λ) distribution is the distribution of $Y = 1/X$ where X has a Gamma(α, λ) distribution. The density is

$$\frac{\lambda^\alpha}{y^{\alpha+1}\Gamma(\alpha)} \mathrm{e}^{-\lambda/y}. \tag{A1.1}$$

The inverse Gamma distribution is popular as Bayesian prior for normal variances (see, e.g., XIII.2.2).

The Inverse Gauss(c, ξ) **distribution** has density

$$\frac{c}{x^{3/2}\sqrt{2\pi}} \exp\left\{\xi c - \frac{1}{2}\left(\frac{c^2}{x} + \xi^2 x\right)\right\}, \quad x > 0. \tag{A1.2}$$

The c.g.f. is $\xi c - c\sqrt{\xi^2 - 2s}$, the mean is c/ξ, and the variance is c/ξ^3. The inverse Gaussian distribution can be interpreted as the time needed for Brownian motion with drift ξ to get from level 0 to level c. The distribution is also popular in statistics as a flexible two-parameter distribution (in fact, one of the nice examples of a two-parameter exponential family). An important extension is the NIG family of distributions; see XII.1.4 and XII.5.1.

The Weibull(β) **distribution** has tail $\overline{F}(x) = \mathrm{e}^{-x^\beta}$. It originates from reliability. The failure rate is ax^b, where $a = \beta$, $b = \beta - 1$, which provides one of the simplest parametric alternatives to a constant failure rate as for the exponential distribution corresponding to $\beta = 1$.

The multivariate normal distribution $\mathscr{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in $\mathbb{R}^d$ has density

$$\frac{1}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\{-(\boldsymbol{x} - \boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2\}.$$

The m.g.f. for $d = 1$ is $\mathrm{e}^{\mu s - s^2\sigma^2/2}$.

An important formula states that conditioning in the multivariate normal distribution again leads to a multivariate normal: if

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim \mathscr{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right),$$

then

$$\boldsymbol{X}_1 \,\big|\, \boldsymbol{X}_2 \sim \mathscr{N}\bigl(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\bigr).$$