Appendix D
Probability Distributions and Generating Functions

Continuous Distributions

Multivariate Normal Distribution

The p-dimensional random variable $\boldsymbol{X} = [X_1, \dots, X_p]^{\mathrm{T}}$ has a normal distribution, denoted $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu} = [\mu_1, \dots, \mu_p]^{\mathrm{T}}$ and $p \times p$ variance–covariance matrix $\boldsymbol{\Sigma}$ if its density is of the form

$$p(\boldsymbol{x}) = (2\pi)^{-p/2} \mid \boldsymbol{\Sigma} \mid^{-1/2} \times \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

for $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{\mu} \in \mathbb{R}^p$ and non-singular $\boldsymbol{\Sigma}$.

Summaries:

$$\begin{aligned} \mathrm{E}[\boldsymbol{X}] &= \boldsymbol{\mu} \\ \mathrm{mode}(\boldsymbol{X}) &= \boldsymbol{\mu} \\ \mathrm{var}(\boldsymbol{X}) &= \boldsymbol{\Sigma}. \end{aligned}$$

Suppose

$$\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \sim \mathrm{N}_p\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{bmatrix}\right)$$

where:

- $\boldsymbol{X}_1$ and $\boldsymbol{\mu}_1$ are $r \times 1$,
- $\boldsymbol{X}_2$ and $\boldsymbol{\mu}_2$ are $(p-r) \times 1$,
- $\boldsymbol{V}_{11}$ is $r \times r$,
- $\boldsymbol{V}_{12}$ is $r \times (p-r)$, $\boldsymbol{V}_{21}$ is $(p-r) \times r$,
- $\boldsymbol{V}_{22}$ is $(p-r) \times (p-r)$.

J. Wakefield, *Bayesian and Frequentist Regression Methods*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0925-1_16,

Then the marginal distribution of $\boldsymbol{X}_1$ is

$$\boldsymbol{X}_1 \sim \mathrm{N}_r(\boldsymbol{\mu}_1, \boldsymbol{V}_{11})$$

and the conditional distribution $\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2$ is

$$\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim \mathrm{N}_r\left[\boldsymbol{\mu}_1 + \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{W}_{11}\right], \tag{D.1}$$

where $\boldsymbol{W}_{11} = \boldsymbol{V}_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{21}$.

Suppose

$$Y_j \mid \mu_j, \sigma_j^2 \sim \mathrm{N}(\mu_j, \sigma_j^2),$$

for $j = 1, \ldots, J$, with $Y_1, \ldots, Y_J$ independent. Then, if $a_1, \ldots, a_J$ represent constants,

$$Z = \sum_{j=1}^{J} a_j Y_j \sim \mathrm{N}\left(\sum_{j=1}^{J} a_j \mu_j, \sum_{j=1}^{J} a_j^2 \sigma_j^2\right). \tag{D.2}$$

If $\boldsymbol{Y}$ is a $p \times 1$ vector of random variables whose distribution is $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A}$ is an $r \times p$ matrix of constants, then

$$\boldsymbol{A}\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}}). \tag{D.3}$$

Beta Distribution

The random variable X follows a beta distribution, denoted $\mathrm{Be}(a, b)$, if its density has the form:

$$p(x) = \mathrm{B}(a, b)^{-1} x^{a-1} (1 - x)^{b-1},$$

for $0 < x < 1$ and $a, b > 0$ and where

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)} = \int_0^1 z^{a-1}(1 - z)^{b-1}\, dz \tag{D.4}$$

is the beta function.

Summaries:

$$\begin{aligned}
\mathrm{E}[X] &= \frac{a}{a + b} \\
\mathrm{mode}(X) &= \frac{a - 1}{a + b - 2} \quad \text{for} \quad a, b > 1 \\
\mathrm{var}(X) &= \frac{ab}{(a + b)^2(a + b + 1)}.
\end{aligned}$$

Gamma Distribution

The random variable X follows a gamma distribution, denoted $\text{Ga}(a, b)$, if its density is of the form

$$p(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx),$$

for $x > 0$ and $a, b > 0$.

Summaries:

$$\begin{aligned} \text{E}[X] &= \frac{a}{b} \\ \text{mode}(X) &= \frac{a-1}{b} \quad \text{for} \quad a \geq 1 \\ \text{var}(X) &= \frac{a}{b^2}. \end{aligned}$$

A χ^2_k random variable with degrees of freedom k corresponds to the $\text{Ga}(k/2, 1/2)$ distribution.

Inverse Gamma Distribution

The random variable X follows an inverse gamma distribution, denoted $\text{InvGa}(a, b)$, if its density is of the form

$$p(x) = \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp(-b/x),$$

for $x > 0$ and $a, b > 0$.

Summaries:

$$\begin{aligned} \text{E}[X] &= \frac{b}{a-1} \quad \text{for} \quad a > 1 \\ \text{mode}(X) &= \frac{b}{a+1} \\ \text{var}(X) &= \frac{b^2}{(a-1)^2(a-2)} \quad \text{for} \quad a > 2. \end{aligned}$$

If Y is $\text{Ga}(a, b)$ then $X = Y^{-1}$ is $\text{InvGa}(a, b)$.

Lognormal Distribution

The random variable X follows a (univariate) lognormal distribution, denoted $\text{LogNorm}(\mu, \sigma^2)$, if its density is of the form

$$p(x) = (2\pi\sigma^2)^{-1/2} \frac{1}{x} \exp\left[-\frac{1}{2\sigma^2}(\log x - \mu)^2\right],$$

for $x > 0$ and $\mu \in \mathbb{R}$, $\sigma > 0$.

Summaries:

$$\begin{aligned} \mathrm{E}[X] &= \exp(\mu + \sigma^2/2) \\ \text{mode}(X) &= \exp(\mu - \sigma^2) \\ \text{var}(X) &= \mathrm{E}[X]^2 \left[\exp(\sigma^2) - 1\right]. \end{aligned}$$

If Y is $\mathrm{N}(\mu, \sigma^2)$ then $X = \exp(Y)$ is $\text{LogNorm}(\mu, \sigma^2)$.

Laplacian Distribution

The random variable X follows a Laplacian distribution, denoted $\text{Lap}(\mu, \phi)$, if its density is of the form

$$p(x) = \frac{1}{2\phi} \exp(- \mid x - \mu \mid /\phi),$$

for $x \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\phi > 0$.

Summaries:

$$\begin{aligned} \mathrm{E}[X] &= \mu \\ \text{mode}(X) &= \mu \\ \text{var}(X) &= 2\phi^2. \end{aligned}$$

Multivariate t Distribution

The p-dimensional random variable $\boldsymbol{X} = [X_1, \ldots, X_p]^{\mathrm{T}}$ has a (Student's) t distribution with d degrees of freedom, location $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_p]^{\mathrm{T}}$ and $p \times p$ scale matrix $\boldsymbol{\Sigma}$, denoted $\mathrm{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, d)$, if its density is of the form

$$p(\boldsymbol{x}) = \frac{\Gamma\left[(d+p)/2\right]}{\Gamma(d/2)(d\pi)^{p/2}} |\boldsymbol{\Sigma}|^{-1/2} \left[1 + \frac{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{d}\right]^{-(d+p)/2},$$

for $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{\mu} \in \mathbb{R}^p$, non-singular $\boldsymbol{\Sigma}$ and $d > 0$.

Summaries:

$$\begin{aligned}
\mathrm{E}[\boldsymbol{X}] &= \boldsymbol{\mu} \quad \text{for} \quad d > 1 \\
\mathrm{mode}(\boldsymbol{X}) &= \boldsymbol{\mu} \\
\mathrm{var}(\boldsymbol{X}) &= \frac{d}{d-2} \times \boldsymbol{\Sigma} \quad \text{for} \quad d > 2.
\end{aligned}$$

The margins of a multivariate t distribution also follow t distributions. For example, if $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]^{\mathrm{T}}$ where $\boldsymbol{X}_1$ is $r \times 1$ and $\boldsymbol{X}_2$ is $(p-r) \times 1$, then the marginal distribution is

$$\boldsymbol{X}_1 \sim \mathrm{T}_r(\boldsymbol{\mu}_1, \boldsymbol{V}_{11}, d),$$

where $\boldsymbol{\mu}_1$ is $r \times 1$ and $\boldsymbol{V}_{11}$ is $r \times r$.

F Distribution

The random variable X follows an F distribution, denoted $\mathrm{F}(a, b)$, if its density is of the form

$$p(x) = \frac{a^{a/2} b^{b/2}}{\mathrm{B}(a/2, b/2)} \frac{x^{a/2-1}}{(b+ax)^{(a+b)/2}},$$

for $x > 0$, with degrees of freedom $a, b > 0$ and where $\mathrm{B}(\cdot, \cdot)$ is the beta function, as defined in (D.4).

Summaries:

$$\begin{aligned}
\mathrm{E}[X] &= \frac{b}{b-2} \quad \text{for} \quad b > 2 \\
\mathrm{mode}(X) &= \frac{a-2}{a} \frac{b}{b+2} \quad \text{for} \quad a > 2 \\
\mathrm{var}(X) &= \frac{2b^2(a+b-2)}{a(b-2)^2(b-4)} \quad \text{for} \quad b > 4.
\end{aligned}$$

Wishart Distribution

The $p \times p$ random matrix $\boldsymbol{X}$ follows a Wishart distribution, denoted $\text{Wish}_p(r, \boldsymbol{S})$, if its probability density function is of the form

$$p(\boldsymbol{x}) = \frac{\mid \boldsymbol{x} \mid^{(r-p-1)/2}}{2^{rp/2}\Gamma_p(r/2) \mid \boldsymbol{S} \mid^{r/2}} \exp\left[-\frac{1}{2}\text{tr}(\boldsymbol{x}\boldsymbol{S}^{-1})\right],$$

for $\boldsymbol{x}$ positive definite, $\boldsymbol{S}$ positive definite and $r > p - 1$ and where

$$\Gamma_p(r/2) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma[(r+1-j)/2]$$

is the generalized gamma function.

Summaries:

$$\begin{aligned} \text{E}[\boldsymbol{X}] &= r\boldsymbol{S} \\ \text{mode}(\boldsymbol{X}) &= (r-p-1)\boldsymbol{S} \quad \text{for} \quad r > p+1 \\ \text{var}(X_{ij}) &= r(S_{ij}^2 + S_{ii}S_{jj}) \quad \text{for} \quad i, j = 1, \ldots, p. \end{aligned}$$

Marginally, the diagonal elements X_{ii} have distribution $\text{Ga}[r/2, 1/(2S_{ii})]$, $i = 1, \ldots, p$.

Taking $p = 1$ yields

$$p(x) = \frac{(2S)^{-r/2}}{\Gamma(r/2)} x^{r/2-1} \exp(-x/2S),$$

for $x > 0$ and $S, r > 0$, i.e. a $\text{Ga}[r/2, 1/(2S)]$ distribution, revealing that the Wishart distribution is a multivariate version of the gamma distribution.

Inverse Wishart Distribution

The $p \times p$ random matrix $\boldsymbol{X}$ follows an inverse Wishart distribution, denoted $\text{InvWish}_p(r, \boldsymbol{S})$, if its probability density function is of the form

$$p(\boldsymbol{x}) = \frac{|\boldsymbol{x}|^{-(r+p+1)/2}}{2^{rp/2}\Gamma_p(r/2) \mid \boldsymbol{S} \mid^{r/2}} \exp\left[-\frac{1}{2}\text{tr}(\boldsymbol{x}^{-1}\boldsymbol{S})\right],$$

for $\boldsymbol{x}$ positive definite, $\boldsymbol{S}$ positive definite and $r > p - 1$.

Summaries:

$$\mathrm{E}[\boldsymbol{X}] = \frac{\boldsymbol{S}^{-1}}{r-p-1} \quad \text{for} \quad r > p+1$$
$$\mathrm{mode}(\boldsymbol{X}) = \frac{\boldsymbol{S}^{-1}}{r+p+1}$$
$$\mathrm{var}(X_{ij}) = \frac{(r-p+1)S_{ij}^{-2} + (r-p-1)S_{ii}^{-1}S_{jj}^{-1}}{(r-p)(r-p-1)^2(r-p-3)} \quad \text{for} \quad i,j = 1,\dots,p.$$

If $p = 1$ we recover the inverse gamma distribution InvGa$[r/2, 1/(2S)]$ with

$$\mathrm{E}[X] = \frac{1}{S(r-2)} \quad \text{for} \quad r > 2$$
$$\mathrm{mode}(X) = \frac{1}{S(r+2)}$$
$$\mathrm{var}(X) = \frac{1}{S^2(r-2)(r-4)} \quad \text{for} \quad r > 4.$$

If $\boldsymbol{Y} \sim$ Wish$_p(r, \boldsymbol{S})$, the distribution of $\boldsymbol{X} = \boldsymbol{Y}^{-1}$ is InvWish$_p(r, \boldsymbol{S})$.

Discrete Distributions

Binomial Distribution

The random variable X has a binomial distribution, denoted Binomial(n, p), if its distribution is of the form

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x},$$

for $x = 0, 1, \dots, n$ and $0 < p < 1$.

Summaries:

$$\mathrm{E}[X] = np$$
$$\mathrm{var}(X) = np(1-p).$$

Poisson Distribution

The random variable X has a Poisson distribution, denoted Poisson(μ), if its distribution is of the form

$$\Pr(X = x) = \frac{\exp(-\mu)\mu^x}{x!},$$

for $\mu > 0$ and $x = 0, 1, 2, \ldots$.

Summaries:

$$\begin{aligned} \mathrm{E}[X] &= \mu \\ \mathrm{var}(X) &= \mu. \end{aligned}$$

Negative Binomial Distribution

The random variable X has a negative binomial distribution, denoted NegBin(μ, b), if its distribution is of the form

$$\Pr(X = x) = \frac{\Gamma(x+b)}{\Gamma(x+1)\Gamma(b)} \left(\frac{\mu}{\mu+b}\right)^x \left(\frac{b}{\mu+b}\right)^b,$$

for $\mu > 0$, $b > 0$ and $x = 0, 1, 2, \ldots$

Summaries:

$$\begin{aligned} \mathrm{E}[X] &= \mu \\ \mathrm{var}(X) &= \mu + \mu^2/b. \end{aligned}$$

The negative binomial distribution arises as a gamma mixture of a Poisson random variable. Specifically, if $X \mid \mu, \delta \sim \text{Poisson}(\mu\delta)$ and $\delta \mid b \sim \text{Ga}(b, b)$, then $X \mid \mu, b \sim \text{NegBin}(\mu, b)$.

We link the above description, motivated by a random effects argument, with the more familiar derivation in which the negative binomial arises as the number of failures seen before we observe b successes from independent trials, each with success probability $p = \mu/(\mu + b)$. The probability distribution is

$$\Pr(X = x) = \binom{x+b-1}{x} p^x (1-p)^b,$$

for $0 < p < 1$, $b > 0$ an integer, and $x = 0, 1, 2, \ldots$

Summaries:

$$\mathrm{E}[X] = \frac{pb}{1-p}$$

$$\mathrm{var}(X) = \frac{pb}{(1-p)^2}.$$

Generating Functions

The *moment generating function* of a random variable Y is defined as $M_Y(t) = \mathrm{E}[\exp(tY)]$, for $t \in \mathbb{R}$, whenever this expectation exists. We state three important and useful properties of moment generating functions:

1. If two distributions have the same moment generating functions then they are identical at almost all points.
2. Using a series expansion:

$$\exp(tY) = 1 + tY + \frac{t^2Y^2}{2!} + \frac{t^3Y^3}{3!} + \dots$$

so that

$$M_Y(t) = 1 + tm_1 + \frac{t^2m_2}{2!} + \frac{t^3m_3}{3!} + \dots$$

where m_i is the ith moment. Hence,

$$\mathrm{E}[Y^i] = M_Y^{(i)}(0) = \left.\frac{d^iM_Y}{dt^n}\right|_{t=0}.$$

3. If $Y_1, \dots, Y_n$ are a sequence of independent random variables and $S = \sum_{i=1}^{n} a_iY_i$, with a_i constant, then the moment generating function of S is

$$M_S(t) = \prod_{i=1}^{n} M_{Y_i}(a_it).$$

The *cumulant generating function* of a random variable Y is defined as

$$C_Y(t) = \log \mathrm{E}[\exp(tY)]$$

for $t \in \mathbb{R}$.