Common Probability Distributions

Uniform $X \sim \mathcal{U}(a, b)$, $a, b \in \mathbb{R}$

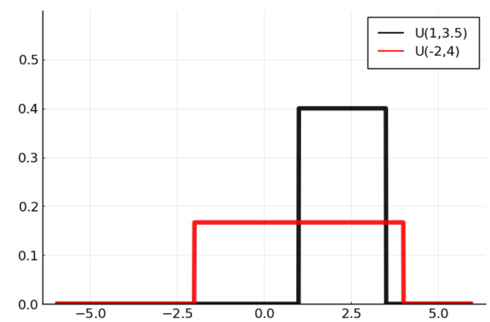

Figure D1: Uniform Distributions

- Domain: $[a, b]$
- pdf: $f(x) = \frac{1}{b-a}$, cdf: $F(x) = \frac{x-a}{b-a}$
- $\mathbb{E}[X] = (a + b)/2$

Normal $X \sim N\left(\mu, \sigma^2\right)$, $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_{++}$

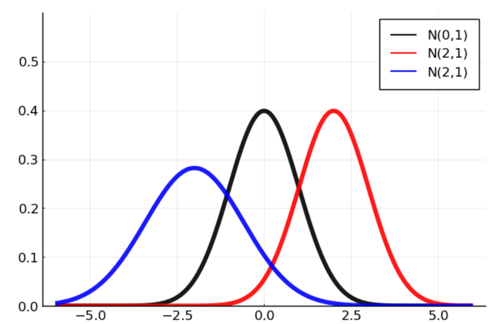

Figure D2: Normal Distributions

- Domain: $\mathbb{R}$
- pdf:

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

- $\mathbb{E}[X] = \mu$, $\mathbb{V}[X] = \sigma^2$, mode $= \mu$

X has the following probabilities of falling within 1, 2 or 3 standard deviations from the mean, respectively:

$$\mathbb{P}\left(\left|\frac{y-\mu}{\sigma}\right| < 1\right) \approx 0.68 \; , \quad \mathbb{P}\left(\left|\frac{y-\mu}{\sigma}\right| < 2\right) \approx 0.95 \; , \quad \mathbb{P}\left(\left|\frac{y-\mu}{\sigma}\right| < 3\right) \approx 0.997 \; .$$

Note that

$$X \sim N\left(\mu, \sigma^2\right) \quad \Leftrightarrow \quad X = \mu + \sigma Z \;\; , \;\; Z \sim N(0, 1) \; .$$

We call $Z = (X - \mu)/\sigma \sim N(0, 1)$ the standardized RV X. We can get the Moment-

Generating Function (MGF) of X based on that of Z. For Z, we have $M_Z(t) = exp\{\frac{1}{2}t^2\}$.[1] For X, we then get $M_X(t) = exp\{\frac{1}{2}\sigma^2 t^2\}exp\{\mu t\}$.

Finally, note that

$$f(x) \propto exp\left\{-\frac{1}{2\sigma^2}(x^2 - 2x\mu)\right\} ,$$

i.e. $f(x) = c\ exp\left\{-\frac{1}{2\sigma^2}(x^2 - 2x\mu)\right\}$. This means that based on this expression we can conclude that $f(x)$ must be the pdf of a Normal distribution because c is a unique constant that is independent of x. It is unique because it makes sure that the expression integrates to one so that $f(x)$ is a valid pdf. It turns out to be $c = (2\pi\sigma^2)^{-\frac{1}{2}}exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}$. In short, if for some RV Y we know $p(y) \propto \exp\left\{-\frac{1}{2b}(y^2 - 2ya)\right\}$, then we can deduce $Y \sim N(a, b)$.[2] [3]

Gamma $X \sim G(\alpha, \beta)$, $\alpha, \beta \in \mathbb{R}_{++}$

- Domain: $\mathbb{R}_{++}$
- pdf:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} exp\{-x/\beta\}$$

- $\mathbb{E}[X] = \alpha\beta$, $\mathbb{V}[X] = \alpha\beta^2$,

$$\text{mode} = \begin{cases} \frac{\alpha-1}{\beta} & \text{for } \alpha > 1 \\ 0 & \text{otherwise} \end{cases}$$

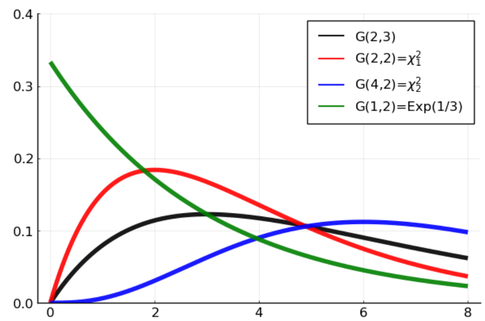

Figure D3: Gamma Distributions

In the pdf-expression, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}exp\{-x\}\,dx$ is the Gamma-function. It has the properties

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha) \quad \text{for } \alpha > 0 , \quad \text{and} \quad \Gamma(n) = (n-1)! \quad \text{for } n \in \mathbb{Z} .$$

The shape parameter α influences rather the peakedness of the distribution, whereas the scale parameter β influences more the variance/spread. Sometimes, a Gamma-distribution is parameterized also with its shape α and rate r, whereby $r = 1/\beta$.

The exponential and chi-squared distributions are special cases of the Gamma distribution:

- Exponential distribution:

$$X \sim \text{Exp}(\lambda) \quad \Leftrightarrow \quad X \sim G(1, 1/\lambda) .^4$$

[1] See Appendix.
[2] See the example of an Exponential distribution in Chapter 1.
[3] Alternatively, if $p(y) \propto \exp\left\{ay^2 + by\right\}$, then we can deduce $Y \sim N\left(-\frac{b}{2a}, -\frac{1}{2a}\right)$.

- Chi-squared distribution with ρ degrees of freedom:

$$X \sim \chi^2_\rho \,, \; \rho \in \mathbb{Z} \quad \Leftrightarrow \quad X \sim G(\frac{\rho}{2}, 2) \,.$$

X has the same distribution as the sum of ρ independent standard Normal RVs, i.e.

$$\{X_i\}_{i=1}^{\rho} \overset{i.i.d.}{\sim} N(0,1) \quad \Rightarrow \quad \sum_{i=1}^{\rho} X_i \sim \chi^2_\rho \,.$$

Also, the Gamma distribution is related to the Inverse Gamma distribution:

$$X \sim G(\alpha, \beta) \quad \Leftrightarrow \quad \frac{1}{X} \sim \text{IG}(\alpha, \beta) \,.$$

Inverse Gamma $\quad X \sim IG(\alpha, \beta) \,, \quad \alpha, \beta \in \mathbb{R}_{++}$

- Domain: $\mathbb{R}_{++}$
- pdf:
$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} exp\{-\beta/x\}$$
- $\mathbb{E}[X] = \beta/(\alpha-1)$ for $\alpha > 1$,
  $\mathbb{V}[X] = \beta^2/(\alpha-1)^2(\alpha-2)$ for $\alpha > 2$,
  mode $= \frac{\beta}{\alpha+1}$

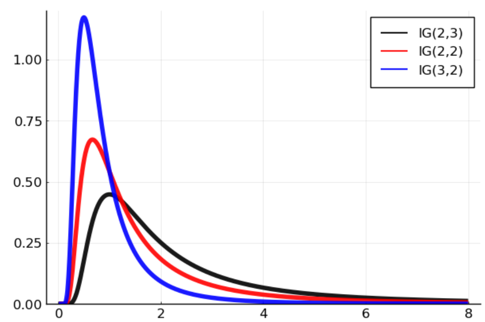

Figure D4: Inverse Gamma Distributions

Sometimes[5] it is useful to reparameterize the Inverse Gamma distribution by writing $X \sim IG(\nu, s^2)$ whereby $\beta = s^2/2$ and $\alpha = \nu/2$. We then get the density

$$f(x) = \frac{s^\nu}{2^{\nu/2}\Gamma(\nu/2)} x^{-(\nu+2)/2} exp\left\{-\frac{s^2}{2x}\right\} \propto x^{-(\nu+2)/2} exp\left\{-\frac{s^2}{2x}\right\} \,.$$

[4] The pdf then simplifies to $f(x) = \lambda exp\{-\lambda x\}$.

[5] e.g. in the context of Bayesian estimation of the error variance σ^2 in a linear regression model (Section 4.5).

Multivariate Normal $X \sim N(\mu, \Sigma)$, $\mu \in \mathbb{R}^k$, $\Sigma_{k \times k}$ positive semi-definite

- Domain: $\mathbb{R}^k$
- pdf:

$$f(x) = (2\pi)^{-k/2}|\Sigma|^{-1/2} exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

- $\mathbb{E}[X] = \mu$, $\mathbb{V}[X] = \Sigma$

First, analogously as for a univariate Normal distribution, we have

$$X \sim N(\mu, \Sigma) \quad \Leftrightarrow \quad X = \mu + \Sigma_{tr} Z \ , \ \ Z \sim N(0, I) \ ,$$

where Σ_{tr} is the Cholesky factor of Σ, i.e. it is a lower-triangular matrix s.t. $\Sigma_{tr}\Sigma_{tr}' = \Sigma$. Hence, based on X, we can get $Z = \Sigma_{tr}^{-1}(X - \mu) \sim N(0, I)$.

Second, let $X \sim N(\mu, \Sigma)$ and partition the vector X into two sub-vectors, $X = (X_1', X_2')'$, and do the corresponding partitions of μ and Σ,

$$\mu = (\mu_1', \mu_2')' \ , \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \ .$$

The marginal pdfs of X_1 and X_2 are then also (multivariate) Normal with the corresponding elements of μ and Σ:

$$X_1 \sim N(\mu_1, \Sigma_{11}) \ , \quad X_2 \sim N(\mu_2, \Sigma_{22}) \ .$$

Note that this implies that every element of X follows a univariate Normal distribution. For the conditional, we get $X_1|X_2 \sim N(\mu_{1|2}, \Sigma_{1|2})$ with

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \ , \quad \Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \ ,$$

and analogously for the pdf of $X_2|X_1$. Based on this result, it turns out that two Normal RVs are independent iff they are uncorrelated, i.e. iff $Cov(X_1, X_2) = \Sigma_{12} = \Sigma_{21}' = 0$ because we can then write the joint pdf of $(X_1', X_2')'$ as a product of the marginal pdfs of X_1 and X_2.

Third,

$$Z \sim N(0, I) \quad \Rightarrow \quad Z'Z =\sim \chi^2_k \ ,$$

because $Z'Z = \sum_{j=1}^k Z_i^2$ is the sum of the k independent $N(0,1)$ RVs $\{Z_i\}_{i=1}^k$ contained in the vector Z. Combining this with the previous result, we have

$$X \sim N(\mu, \Sigma) \quad \Rightarrow \quad (X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2_k \ .$$

Fourth, note that under $\Sigma = \sigma^2 V$ the pdf of $X \sim N(\mu, \Sigma)$ can be written in two ways:

$$\begin{aligned} f(x) &= (2\pi)^{-k/2}|\sigma^2 V|^{-1/2} exp\left\{-\frac{1}{2}(x-\mu)'\left[\sigma^2 V\right]^{-1}(x-\mu)\right\} \\ &= (2\pi\sigma^2)^{-k/2}|V|^{-1/2} exp\left\{-\frac{1}{2\sigma^2}(x-\mu)'V^{-1}(x-\mu)\right\} .^6 \end{aligned}$$

Finally, and analogously as for the univariate Normal distribution, note that

$$f(x) \propto exp\left\{-\frac{1}{2}x'\Sigma^{-1}x - 2\mu'\Sigma^{-1}x\right\} .^7$$

Therefore, if for some RV Y we have $p(y) \propto exp\left\{-\frac{1}{2}x'Ax - 2b'Ax\right\}$, then we can deduce $Y \sim N\left(b, A^{-1}\right)$.[8]

Matrix-Variate Normal $X \sim MN(\mu, U, V)$, $\mu_{n\times k} \in \mathbb{R}^{nk}$, $U_{n\times n}, V_{k\times k}$ p.s.d.

- Domain: $X \in \mathbb{R}^{nk}$ is an $n \times k$ matrix
- pdf:

$$f(x) = (2\pi)^{-nk/2}|V|^{-n/2}|U|^{-p/2} exp\left\{-\frac{1}{2}tr\left[V^{-1}(X-\mu)'U^{-1}(X-\mu)\right]\right\}$$

- $\mathbb{E}[X] = \mu$

U determines the variance among rows of X (which gets scaled by the trace (sum of diagonal elements) of V) and V determines the variance among columns of X (which gets scaled by the trace of U):

$$\mathbb{E}[(X-\mu)(X-\mu)'] = U tr[V] \ , \quad \mathbb{E}[(X-\mu)'(X-\mu)] = V tr[U] \ .$$

We can write a Matrix-Normal distribution for the matrix X equivalently as a multivariate Normal distribution for the vectorized X, $vec(X)$, obtained by stacking all columns of X on

[6]This follows from the facts that, for a scalar λ and a $(k \times k)$ matrix A, we have $|\lambda A| = \lambda^k |A|$ and $[\lambda A]^{-1} = \lambda^{-1}A^{-1}$.

[7]Note that Σ is symmetric, and $x'\Sigma^{-1}\mu = \mu'\Sigma^{-1}x$ is a scalar.

[8]Alternatively, if $p(y) \propto \exp\{y'Ay + b'y\}$, then we can deduce $Y \sim N\left(-\frac{1}{2}A^{-1}b, -\frac{1}{2}A^{-1}\right)$.

top of each other into a vector:

$$X \sim MN(\mu, U, V) \quad \Leftrightarrow \quad vec(X) \sim N(vec(\mu), V \otimes U) \ ,$$

i.e. one can rewrite $f(x)$ above as

$$f(x) = (2\pi)^{-nk/2}|V \otimes U|^{-1/2} exp\left\{ -\frac{1}{2}(\overrightarrow{X} - \overrightarrow{\mu})'[V \otimes U]^{-1}(\overrightarrow{X} - \overrightarrow{\mu}) \right\} \ ,$$

where $\overrightarrow{X} = vec(X)$.

Inverse Wishart $X \sim IW(\Psi, \nu)$, $\Psi_{k\times k}$ p.d., $\nu > k-1$, $\nu \in \mathbb{R}$

- Domain: X is a $k \times k$ p.d. matrix
- pdf:
$$f(x) = \frac{|\Psi|^{\nu/2}}{2^{\nu k/2}\Gamma_k(\frac{\nu}{2})}|X|^{-(\nu+k+1)/2} exp\left\{ -\frac{1}{2} tr[\Psi X^{-1}] \right\}$$
- $\mathbb{E}[X] = \frac{\Psi}{\nu-k-1}$ for $\nu > k+1$,
  mode $= \frac{\Psi}{\nu+k+1}$

Appendix

Claim. *For $Z \sim N(0,1)$, we have the MGF $M_Z(t) = exp\{\frac{1}{2}t^2\}$.*

Proof: We have

$$\begin{aligned} M_Z(t) &= \mathbb{E}[exp\{tz\}] \\ &= \int exp\{tz\} \frac{1}{\sqrt{2\pi}} exp\left\{ -\frac{1}{2}z^2 \right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \int exp\left\{ -\frac{1}{2}\left(z^2 - 2t\right) \right\} dz \\ &= exp\left\{ \frac{1}{2}t^2 \right\} \int \frac{1}{\sqrt{2\pi}} exp\left\{ -\frac{1}{2}(z-t)^2 \right\} dz \\ &= exp\left\{ \frac{1}{2}t^2 \right\} \ , \end{aligned}$$

because the expression inside the integral is the pdf of a $N(t,1)$ and therefore has to integrate to one. ■