Analogously, the buoyancy term is parameterized as

$$R_{ij}^{(b)} = -C_b \left(G_{ij} - \frac{2}{3}\delta_{ij} G \right) \tag{101}$$

where

$$G \equiv -\frac{g_i}{\theta_{ref}} \overline{u_i' \theta_v'} \tag{102}$$

and

$$G_{ij} \equiv -\frac{1}{\theta_{ref}} \left(g_j \overline{u_i' \theta_v'} + g_i \overline{u_j' \theta_v'} \right) \tag{103}$$

0.3 A few distributions used in the atmospheric sciences

Before considering the functional form of various idealized probability density functions (PDFs), let us first review what a PDF is. A PDF may be interpreted as follows: $P(x)\ dx$ is the probability of choosing a value between x and $x + dx$. $P(x)$ has inverse units to those of dx, and so $P(x)\ dx$ is dimensionless.

0.3.1 Univariate uniform distribution

A uniform distribution is not the highest accuracy description of small-scale variability in the atmosphere, but it is important because standard random number generators produce numbers that are uniformly distributed on the range $(0, 1)$. Such numbers are equally likely to be found anywhere in the range $(0, 1)$.

Draw sketch of uniform PDF.

0.3.2 Univariate normal distribution

A univariate normal with mean μ and standard deviation σ can be written:

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \tag{104}$$

It extends from $-\infty$ to ∞ and has a symmetric "bell" shape with a single maximum. Sometimes a normal distribution is called a "Gaussian" distribution.

Draw a sketch of a normal PDF.

0.3.3 Multivariate normal distribution

Section 4.1, Johnson (1987), Section 7.4, Press et al. (2007)

Consider a m-dimensional random vector $\mathbf{X} = (X_1, ..., X_m)^T$. Here, $\mathbf{X}$ is a column vector and $()^T$ denotes the transpose. Suppose the distribution of $\mathbf{X}$ is a m-dimensional multivariate normal (Gaussian) distribution. Denote it $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here, $\boldsymbol{\mu}$ is the mean (expection) of vector $\mathbf{X}$, and $\boldsymbol{\Sigma}$ is a $m \times m$ covariance matrix, where $\boldsymbol{\Sigma}_{ij} = \mathrm{Cov}(x_i, x_j)$. $\mathcal{N}_m$ is given by:

$$\mathcal{N}_m(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{m/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\mathbf{x} - \boldsymbol{\mu})\right] \tag{105}$$

Here det denotes the determinant.

For instance, consider a bivariate (i.e. 2D) normal distribution with covariance matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & r_{12}\sigma_x\sigma_y \\ r_{12}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}. \tag{106}$$

Then

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{(1 - r_{12}^2)} \begin{bmatrix} 1/\sigma_x^2 & -r_{12}/(\sigma_x\sigma_y) \\ -r_{12}/(\sigma_x\sigma_y) & 1/\sigma_y^2 \end{bmatrix}. \tag{107}$$

The PDF has the form

$$\mathcal{N}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - r_{12}^2}} \exp\left\{-\frac{1}{2}\frac{1}{1 - r_{12}^2}\left[\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} - \frac{2r_{12}(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y}\right]\right\}, \tag{108}$$

where r_{12} is the correlation between x and y.

A conditional distribution is the distribution that results when one or more variates are held fixed. Suppose that the set of variates $\mathbf{X}$ is divided into two groups:

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \tag{109}$$

with $\mathbf{X}_1$ having dimension k and $\mathbf{X}_2$ having dimension $m - k$. The vector of means may be similarly split:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}. \tag{110}$$

So may the covariance matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}, \tag{111}$$

with dimensions

$$\mathbf{\Sigma} = \begin{bmatrix} k \times k & k \times (m-k) \\ (m-k) \times k & (m-k) \times (m-k) \end{bmatrix}. \tag{112}$$

The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2 = \mathbf{x}_2$ is also a normal distribution. Draw sketch of conditional distribution. Let's write down the conditional distribution without derivation. The conditional mean is given by

$$E(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) = \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2). \tag{113}$$

So consider the example of a 2x2 matrix. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are positively correlated, then $\mathbf{\Sigma}_{12} > 0$. So $E(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) > \boldsymbol{\mu}_1$ if $\mathbf{x}_2 > \boldsymbol{\mu}_2$.

The covariance matrix of the conditional distribution is given by

$$\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \tag{114}$$

If $\mathbf{X}_1$ and $\mathbf{X}_2$ are uncorrelated, then the conditional variance equals the marginal variance. But if $\mathbf{X}_1$ and $\mathbf{X}_2$ have either positive or negative correlation, then the conditional variance is less. The part of the distribution with a defined value of $\mathbf{X}_2$ is narrower than the overall distribution. The conditional distribution does not exist if $\mathbf{\Sigma}_{22}$ is singular.

The conditional distribution itself is given by

$$\mathcal{N}_k(\mathbf{x}_1|\boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}) \tag{115}$$

0.3.4 Mixture of two normal distributions

Section 4.2, Johnson (1987)

$$M_m(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = p\mathcal{N}_m(\mathbf{x}|\boldsymbol{\mu}_1, \mathbf{\Sigma}_1) + (1-p)\mathcal{N}_m(\mathbf{x}|\boldsymbol{\mu}_2, \mathbf{\Sigma}_2) \tag{116}$$

Here p is the mixture fraction, with $0 < p < 1$. p is the probability of choosing component 1, and $1 - p$ is the probability of choosing component 2.

A normal mixture is useful because it can take on either positive or negative skewness, or it can reduce to a single normal, or it can reduce to a delta function plus a normal.

Show some example plots.

0.3.5 Lognormal distribution

Equation 5.1, Johnson (1987); Section 5.2.1, Gentle (2003)

A lognormal is the distribution of a random variable whose logarithm is normally distributed. Thus, if X is lognormally distributed, then $Y = \ln X$ is normally distributed. Conversely, if Y is normally distributed, then $X = \exp(Y)$ is lognormally distributed.

One can "derive" a lognormal distribution by starting from a normal distribution for $\ln x$:

$$\mathcal{N}(\ln x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]. \tag{117}$$

(Note that the dimensions of this equation do not make sense unless μ and σ are dimensionless.) The transformation from this normal distribution to a lognormal distribution must obey the fundamental transformation law of probability. For any transformation of variables $y = y(x)$, then the corresponding distributions $P(x)$ and $P(y)$ are related by (Press et al. 2007)

$$|P(y)dy| = |P(x)dx|\,. \tag{118}$$

To understand this formula in a more graphical way, think of $P(y)$ as a histogram with many bins (rectangles), each made of clay. The height, $P(y)$, and width dy of each bin is stretched, preserving the area (probability) of the bin. In our case, $y = \ln x$, and

$$\mathcal{N}(y)\, dy = P_L(x)\, dx, \tag{119}$$

where P_L is lognormally distributed. In other words,

$$\mathcal{N}(\ln x)\, d\ln x = P_L(x)\, dx, \tag{120}$$

Letting $d\ln x = dx/x$, we find that

$$P_L(x) = \mathcal{N}(\ln x)/x. \tag{121}$$

Therefore, $P_L(x)$ has the form

$$P_L(x) = \frac{1}{(2\pi)^{1/2}\sigma x} \exp\left[\frac{-(\ln x - \mu)^2}{2\sigma^2}\right]. \tag{122}$$

Here μ and σ^2 are the mean and the variance, respectively, of $\ln x$, and $\ln x$ is normally distributed.

A lognormal distribution is useful for representing variables that are non-negative. A lognormal distribution extends from 0 to ∞. Lognormal distributions are positively skewed and have very long tails to the right. To see this, recall that if Y is distributed according to a standard normal, then $\exp(Y)$ is lognormally distributed. All the sample points in the left half of the standard normal distribution Y will be mapped into the region between 0 and 1 in the lognormal distribution $\exp(Y)$. A sample point on the right tail of Y will be mapped to a point exponentially far to the right. This leads to a distribution that is highly skewed to the right.

Sketch a lognormal distribution.

0.4 One way to draw samples from a univariate distribution: The inverse distribution method

Section 2.1, Johnson (1987)

There are several ways to draw a sample point from a univariate distribution. Here, we discuss the inverse distribution method.

Suppose that the probability density function (PDF) of a single variable within a grid box has been assumed to have a particular shape. Now we wish to draw samples from that PDF. The problem is that random number generators typically draw samples from a uniform distribution between $(0, 1)$, whereas we may wish to draw samples that are distributed according to a different PDF, such as a normal PDF. Suppose the PDF is associated with a cumulative distribution function (CDF) F, where

$$F(x) = \text{Prob}(X \leq x) \tag{123}$$

or, in other words,

$$F(x) = \int_{-\infty}^{x} P(x')dx'. \tag{124}$$

If $P(x')$ is non-zero over the full domain, then $F(x)$ is strictly increasing. For instance, the CDF of a standard normal PDF is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-x'^2/2)dx', \tag{125}$$

which is related to the error function by

$$F(x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right]. \tag{126}$$

If F is strictly increasing, a sample X can be drawn from $P(x)$ via the following steps:

1. Generate a variate U that is distributed uniformly between $(0, 1)$.
2. Let $X = F^{-1}(U)$.

Sketch picture of sample point being transformed from a uniform distribution to a standard normal distribution.

Show how the density of sample points changes when we draw from a normal rather than a uniform distribution.

0.5 One way to draw sample points from a multivariate PDF

Section 4.1, Johnson (1987)

0.5.1 Normally distributed sample points

There are several methods to generate variates. Here, we present a three-step method that generates samples distributed according to a multivariate normal with dimension m:

1. *Draw a sample point,* $\mathbf{U}$*, from a uniform (0,1) distribution.* Assume that $\mathbf{U}$ is multivariate, with dimension m. $\mathbf{U}$ can be thought of as a m-dimensional column vector.
2. *Transform* $\mathbf{U}$ *to a sample* $\mathbf{Y}$ *distributed according to a standard normal,* $\mathcal{N}_m(\mathbf{0}, I)$. By definition, a standard normal has zero mean, unit variance, and a covariance matrix equal to the identity matrix, I. Because the variates of $\mathcal{N}_m(\mathbf{0}, I)$ are uncorrelated with each other, each variate of $\mathbf{U}$ can be independently transformed to a variate of $\mathbf{Y}$, using, e.g., the inverse distribution method described above.
3. *Transform the standard normal sample* $\mathbf{Y}$ *to a sample* $\mathbf{X}$ *distributed according to the desired normal distribution,* $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*, which has the desired mean,* $\boldsymbol{\mu}$*, and covariance,* $\boldsymbol{\Sigma}$. To transform a standard normal to a non-standard normal, we distort the sample points so that they have the desired covariance and center the sample points about the desired mean. One way to do this uses the Cholesky decomposition, which factors $\boldsymbol{\Sigma}$ into two matrices: $\mathbf{L}\mathbf{L}^T = \boldsymbol{\Sigma}$. Then we may compute

$$\mathbf{X} = \mathbf{L} \cdot \mathbf{Y} + \boldsymbol{\mu} \tag{127}$$

For example, if the covariance matrix is given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}, \tag{128}$$

then the Cholesky matrix $\mathbf{L}$ is given by

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ r_{12} & \sqrt{1 - r_{12}^2} \end{bmatrix}. \tag{129}$$

Now suppose, for example, that the desired PDF is uncorrelated, with $r_{12} = 0.0$ and $\boldsymbol{\mu} = 0$. Then $\mathbf{L}$ is the identity matrix and Eqn. (127) leaves the sample points $\mathbf{Y}$ unchanged: $\mathbf{X} = \mathbf{Y}$. At the opposite extreme, suppose that the desired PDF is highly correlated, with $r_{12} = 0.99$. Then we can see that $\mathbf{L} \cdot \mathbf{Y}$ will yield highly correlated samples $\mathbf{X}$.

0.5.2 Sample points distributed according to a normal mixture

Section 4.2, Johnson (1987)

Consider the normal mixture

$$pN_m(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) + (1-p)N_m(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2). \tag{130}$$

To generate a sample point from this normal mixture, an extra random number must be generated:

1. Generate a univariate, uniformly distribution point U on $(0, 1)$.
2. If $U \leq p$, generate $\mathbf{X}$ as $N_m(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$.
3. If $U > p$, generate $\mathbf{X}$ as $N_m(\mathbf{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$.

0.5.3 Lognormally distributed sample points

Section 5.2.1, Gentle (2003)

In order to generate a sample point $\mathbf{X_L}$ from a lognormal distribution, we first generate a sample $\mathbf{X}$ from a multivariate normal, $N_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then we exponentiate the sample point:

$$\mathbf{X_L} = \exp(\mathbf{X}) \tag{131}$$

0.6 Basic Monte Carlo integration

Chapter 2, Owen (2013); Section 7.1, Gentle (2003); Section 7.7, Press et al. (2007); Sections 2.6–2.7, Kalos and Whitlock (2008)

Suppose we wish to average a function, $f(x)$, multiplied by a PDF, $P(x)$, over a domain D:

$$\overline{f} = \int_D f(x)P(x)\,dx. \tag{132}$$

This may be thought of as a weighted sum: the values of $f(x)$ are weighted by the weights $P(x)\,dx$ and then summed. Note that Eqn. (132) reduces to an ordinary integral $\int_D f(x)dx$ if $P(x)$ is a uniform distribution. The integral in Eqn. (132) may also be interpreted as the area under a curve, $f(x)P(x)$.

Integrals such as Eqn. (132) are often approximated by deterministic quadrature methods that discretize the domain and sum the contributions from each part of the domain. Monte Carlo integration is a different approach. It has a more statistical flavor. We view the problem of computing the integral $\overline{f}$ in (132) as the problem of estimating the expectation of a function of a random variable: $\overline{f} = \mathbb{E}[f(X)]$ with distribution $P(x)$. Simple Monte Carlo integration has the following steps:

1. Draw a set of N sample points, x_i, from the PDF.
2. Evaluate $f(x_i)$ at each sample point.
3. Sum the values of f:

$$\overline{f} \approx \hat{f}_N \equiv \frac{1}{N}\sum_{i=1}^{N} f(x_i). \tag{133}$$

How much error is involved in this approach? Consider a random number X and its Monte Carlo estimate, $\hat{\mu}_N$. First let us show that $\hat{\mu}_N$ is unbiased. Let $\mu \equiv \mathbb{E}(X)$, and assume that the N samples are independent and identically distributed. Then,

$$\mathbb{E}(\hat{\mu}_N) = \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(X_i) = \frac{1}{N}\sum_{i=1}^{N}\mu = \frac{1}{N}N\mu = \mu. \tag{134}$$

How much spread can we expect in independent calculations of $\hat{\mu}_N$? Let's calculate the variance of

the estimator $\hat{\mu}_N$, $\mathrm{Var}(\hat{\mu}_N)$:

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_N) &= \mathbb{E}\left[(\hat{\mu}_N - \mu)^2\right] \\
&= \mathbb{E}\left[\left(\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) - \mu\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} (X_i - \mu)\right)^2\right] \\
&\equiv \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} X_i'\right)^2\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{N} X_i'\right)^2\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{N} X_i'\right)\left(\sum_{j=1}^{N} X_j'\right)\right] \\
&= \frac{1}{N^2}\mathbb{E}\left[\sum_{i=1}^{N} X_i'^2 + \sum_{i=1}^{N}\sum_{j\neq i}^{N-1} X_i' X_j'\right] \\
&= \frac{1}{N^2}\left[\sum_{i=1}^{N} \mathbb{E}\left(X_i'^2\right) + \sum_{i=1}^{N}\sum_{j\neq i}^{N-1} \mathbb{E}\left(X_i' X_j'\right)\right].
\end{aligned} \tag{135}$$

If the points are IID distributed, then the second term on the right-hand side is 0. But since $X_i' \equiv X_i - \mu$, then

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_N) &= \frac{1}{N^2}\sum_{i=1}^{N} \mathbb{E}\left[(X_i - \mu)^2\right] \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}\left(X_i\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \sigma^2 \\
&= \frac{1}{N^2} N\sigma^2 \\
&= \frac{1}{N}\sigma^2.
\end{aligned} \tag{136}$$

Therefore, the standard deviation of $\hat{\mu}_N$ is $\sigma/\sqrt{N}$.

Although the standard deviation does give us an overall estimate of the error in $\hat{\mu}_N$, it does not give us confidence intervals. E.g., it does not tell us the range of values within which $\hat{\mu}_N$ lies with, say, a 95% confidence. For that, we need information about the tails of the distribution of $\hat{\mu}_N$.

Fortunately, the central limit theorem states that, in the limit of a large number of samples N, the distribution of $\hat{\mu}_N$ is a normal distribution. The normal distribution has a mean of μ and a variance

of σ/N. It is given by

$$\frac{1}{\sqrt{2\pi\left(\sigma^2/N\right)}}\exp\left[-\frac{1}{2}\frac{\left(\hat{\mu}_N-\mu\right)^2}{\left(\sigma^2/N\right)}\right]. \tag{137}$$

Here $\sigma^2/N = \mathrm{Var}\left(\hat{\mu}_N\right)$. A derivation of this is given in Section 2.7 of Kalos and Whitlock (2008) and Section 2.2 of Owen (2013).

Given the distribution, a 95% confidence interval may be formed by finding the left and right edge of the middle 95% of the area under the distribution. The area is given by the cumulative distribution function (CDF) of the distribution. Consider first a standard normal distribution with CDF Φ. The left edge of the right 2.5% tail is given by

$$\Phi^{-1}(1-0.025) \approx 1.96. \tag{138}$$

The right edge of the left 2.5% tail is given by

$$\Phi^{-1}(0.025) \approx -1.96. \tag{139}$$

Using the fact that a nonstandard normal has a mean of μ and a standard deviation of $\sigma/\sqrt{N}$, we find that a 95% confidence interval for $\hat{\mu}_N$ is

$$\hat{\mu}_N - 1.96\frac{\sigma}{\sqrt{N}} \leq \mu \leq \hat{\mu}_N + 1.96\frac{\sigma}{\sqrt{N}}. \tag{140}$$

In these estimates, the distribution standard deviation, σ, may be replaced by the sample estimate of σ, s, where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{N}(X_i - \hat{\mu}_N)^2. \tag{141}$$

The Monte Carlo method has several advantages:

1. Monte Carlo integration is non-intrusive. Suppose that $f(x)$ is a complicated numerical subroutine created by a specialist. Intruding upon the subroutine's code to modify it or computing the derivative of $f(x)$ might be labor intensive. But Monte Carlo integration requires no modification or extension of $f(x)$.

2. Conventional quadrature methods like Simpson's rule may perform poorly for non-smooth integrands. However, Monte Carlo integration is not degraded for functions that are not smooth.

3. The convergence rate of Monte Carlo integration is independent of the dimensionality of the integral, m. The variance of the integration estimate, σ^2/N, does not contain m. In contrast, a typical quadrature method has a convergence rate of $\mathrm{O}(N^{-4/m})$. For this reason, Monte Carlo integration is an especially useful method for high-dimensional integrals.

4. The error $\sigma/\sqrt{N}$ of Monte Carlo integration can be estimated, thereby allowing the user to adjust the number of sample points until a desired error tolerance is matched.