Mathematical Modeling of the Atmosphere Summary

Sven Bergmann

January 22, 2023

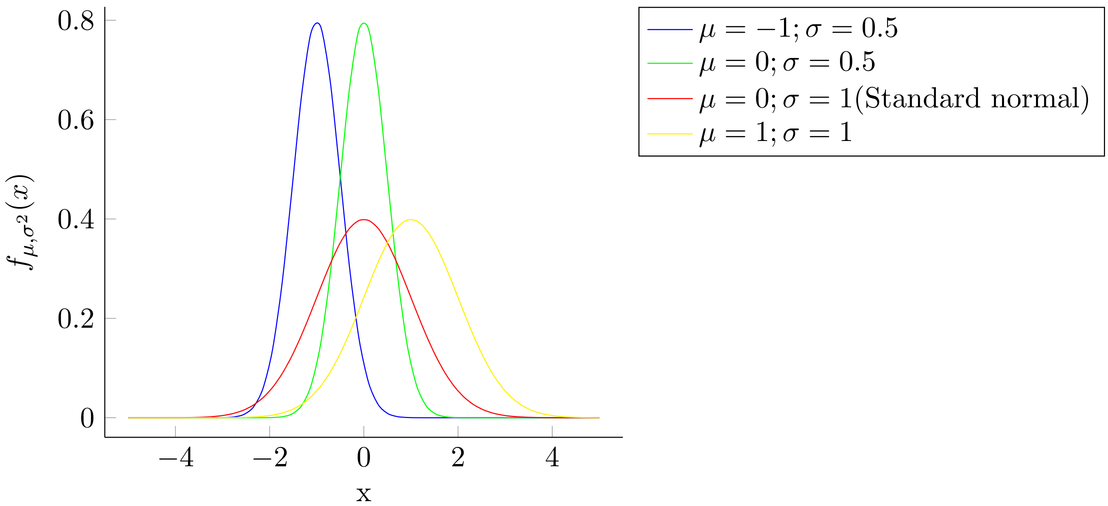
Contents

1 Important distributions

1.1 Normal Distribution

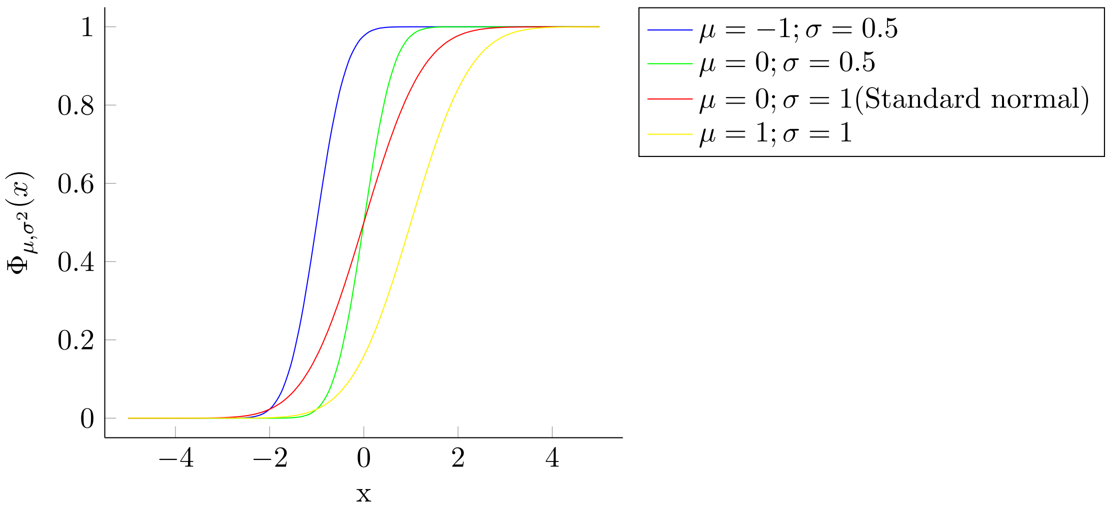
Probability density function:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$



Cumulative distribution function:

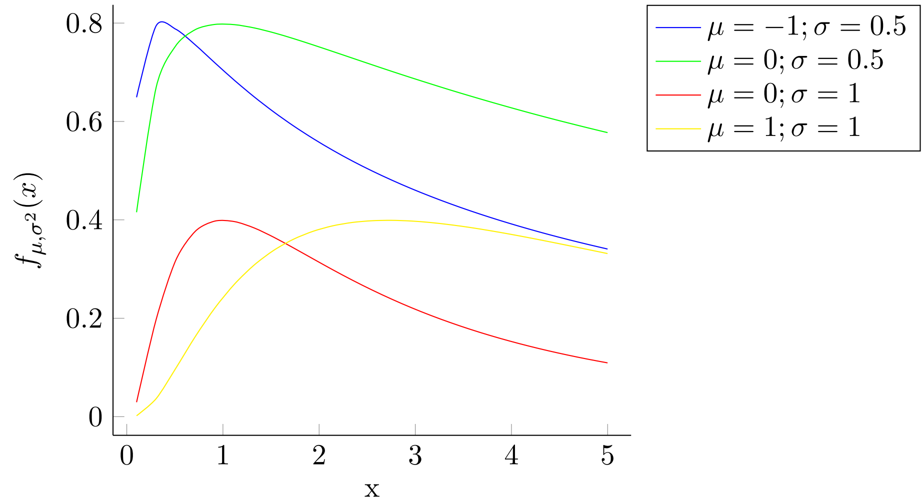
$$\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$



1.2 Lognormal Distribution

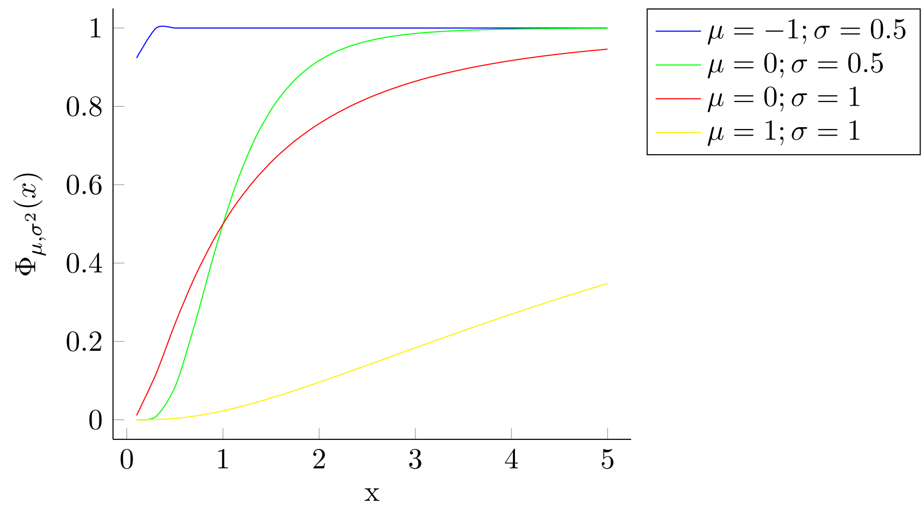
Probability density function:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$



Cumulative distribution function:

$$F(x) = \Phi(\frac{\ln(x) - \mu}{\sigma})$$



1.3 Standard Bivariate Normal Distribution

2 Reynolds Rules

For any variable, the observed value can be written as a sum of the mean value and a turbulent value. The turbulent part is sometimes called the fluctuating part or the perturbation. For example, we can write the zonal component of the velocity as:

$$u = \overline{u} + u'$$

where $\overline{u}$ is the mean, or average value, and u' is the turbulent part.

The average can be a spatial average or a temporal average. Mathematically, we can write those averages using integration if we have continuous data or by summation if we have discrete data. Here are the formulas for computing the mean of the zonal component of the velocity:

Temporal Averaging:

$$\overline{u} = \frac{\int u(t)dt}{\int dt} = \frac{\sum\limits_{i=0}^{N-1} u(t_i)}{N}$$

Spatial Averaging:

$$\overline{u} = \frac{\int u(x)dx}{\int dx} = \frac{\sum\limits_{i=0}^{N-1} u(x_i)}{N}$$

where the integrals are over specific time periods or distances and the summations are for N discrete measurements. If the turbulence does not change with time and is homogeneous (i.e., the same in all directions and for all time), then these averages equal each other. It is important to remember that the mean values defined in the above equation are not necessarily constant in space or time. However, the variations of the mean values will generally be much smaller than the variations of the turbulent values.

In Lesson 10, we developed the equation of motion without really considering the short-term and small-scale variations, except to say that they led to turbulent drag, which acts to resist the mean flow in the upper boundary layer. Now we want to think about how to correctly capture the dynamic effects of turbulent motion. What we want to do is to write down the equations of motion that you learned in Lesson 10; substitute mean and turbulent parts for the variables such as u, v, and w; average over all the terms; and then see if we can sort out the terms to create an equation for the mean wind and an equation for the turbulent wind. This type of averaging is called Reynolds averaging.

But first we need to learn the rules for averaging.

2.1 Rules of Averaging

c is constant; u and v are variables

$$\overline{u'} = 0$$

$$\overline{cu} = c\overline{u}$$

$$\overline{u+v} = \overline{u} + \overline{v}$$

$$\overline{(\overline{u}v)} = \overline{uv}$$

$$\overline{(\frac{\delta u}{\delta t})} = \frac{\delta \overline{u}}{\delta t}$$

3 Variance Reduction

Improves the efficiency of Monte Carlo methods. Problem:

$$\text{Find } \mu = \mathbb{E}(f(X)) \Leftrightarrow \mu = \int_D f(X)P(X)dx, D \subset \mathbb{R}^d$$

3.1 Antithetic sampling

The antithetic variates technique consists, for every sample path obtained, in taking its antithetic path – that is given a path $\{\varepsilon_1, \ldots, \varepsilon_M\}$ to also take $\{-\varepsilon_1, \ldots, -\varepsilon_M\}$. The advantage of this technique is twofold: it reduces the number of normal samples to be taken to generate N paths, and it reduces the variance of the sample paths, improving the precision.

$$\hat{\mu}_{anti} = \frac{1}{n}\sum_{i=1}^{n/2}(f(X_i) + f(\tilde{X}_i))$$

with n as even number, $\tilde{X}_i \overset{iid}{\sim} P$, $\tilde{X} = -X$ for $P \sim \mathcal{N}(0, \Sigma)$, $\tilde{X} = 1 - X$ for $P \sim \mathcal{U}(0,1)^d$

$$\begin{aligned} Var(\hat{\mu}_{anti}) &= \frac{1}{2n}(Var(f(X)) + Var(f(\tilde{X})) + 2 \cdot Cov(f(X), f(\tilde{X})) \\ &= \frac{\sigma^2}{n}(1-p) \end{aligned}$$

$$\implies -1 \leq p \leq 1 : 0 \leq \sigma^2(1-p) \leq 2\sigma^2$$

3.2 Stratification

Split $\mathcal{D}$ into seperate regions called strata.

$$\hat{\mu}_{strata} = \sum_{j=1}^{J} \frac{\omega_j}{n_j} \sum_{i=1}^{n_j} f(X_{ij})$$

$$Var(\hat{\mu}_{strata}) = \sum_{j=1}^{J} \omega_j^2 \frac{\omega_j^2}{n_j}$$

$$\sigma^2 = \underbrace{\sum_{j=1}^{J} \omega_j \sigma_j^2}_{\text{within domain variance}} + \underbrace{\sum_{j=1}^{J} \omega_j (\mu_j - \mu)^2}_{\text{between domain variance}}$$

3.3 Control variates

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(X_i), \hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

Difference estimator:

$$\hat{\mu}_{\text{diff}} = \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - h(X_i)) + \Theta = \hat{\mu} - \Theta + \Theta$$

Regression estimator:

$$\begin{aligned}
\hat{\mu}_{diff} &= \frac{1}{n}\sum_{i=1}^{n}(f(X_i) - \beta h(X_i)) + \beta\Theta = \hat{\mu} - \beta(\hat{\Theta} - \Theta) \\
Var(\hat{\mu}_{\beta}) &= \frac{1}{n}(Var(f(X))) - 2\beta Cov(f(X), h(X)) + \beta^2 Var(h(X))) \\
\beta_{opt} &= \frac{Cov(f(X), h(X))}{Var(h(X))} = \frac{\mathbb{E}((h(X) - \Theta)f(X)))}{\mathbb{E}((h(X) - \Theta))^2} \\
Var(\hat{\mu}_{\beta_{opt}}) &= \frac{\sigma^2}{n}(1 - p^2)
\end{aligned}$$

(In practice:)

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(f(X_i) - \overline{f})(h(X_i) - \overline{h})}{\sum_{i=1}^{n}(h(X_i) - \overline{h})^2}$$

3.4 Basic importance sampling

$$\begin{aligned}
\hat{\mu}_q &= \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)P(X_i)}{q(X_i)} = \int \frac{f(X)P(X)}{q(X)}q(X)dx, X_i \sim q. \\
\sigma_q^2 &= \frac{1}{n}\sum_{i=1}^{n}(\frac{f(X_i)P(X_i)}{q(X_i)} - \hat{\mu}_q)^2 = \frac{1}{n}\sum_{i=1}^{n}(\omega_i f(X_i) - \hat{\mu}_q)^2
\end{aligned}$$

3.5 Control variates in importance sampling

$$\begin{aligned}
\hat{\mu}_{q,\beta} &= \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)P(X_i) - \beta^T h(X_i)}{q(X_i)} + \beta^T\Theta, X_i \overset{iid}{\sim} q \\
\hat{\sigma}_{q,\beta}^2 &= \frac{1}{n - J - 1}\sum_{i=1}^{n}(Y_i - \hat{\mu}_{q,\hat{\beta}} - \hat{\beta}^T Z_i)^2 \\
\hat{\beta} &= \beta_{opt} + O_p(n^{-\frac{1}{2}}) \\
\hat{\mu}_{q,\hat{\beta}} &= \hat{\mu}_{q,\beta_{opt}} + O_p(n^{-1})
\end{aligned}$$

3.6 Mixture importance sampling

$$\hat{\mu}_\alpha = \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i)P(X_i)}{\sum_{j=1}^{J}\alpha_j q_j(X_i)}$$

3.6.1 Special case: defensive importance sampling

$$q_\alpha(X) = \alpha_1 P(X) + \alpha_2 q(X), \alpha_1 + \alpha_2 = 1$$
$$Var(\hat{\mu}_{q,\alpha}) = \frac{1}{n\alpha_1}(\sigma_p^2 + \alpha_2\mu^2)$$

4 Problem Sets

4.1 Problem Set 1: Spatial filtering

Consider a 1D field

$$f(x'') = \frac{1}{2}(1 - \cos(x''))$$

that extends indefinitely to the left and the right. In this problem, we'll see how running-mean filtering of f affects the perturbations of f, namely f'', from the filtered version of f, namely $\overline{f}$. Assume that the filter used is a simple box (uniform) filter, extending from $x'' = x - \frac{\pi}{2}$ to $x'' = x + \frac{\pi}{2}$. Throughout this problem, when you are asked to plot functions, you may sketch the functions by hand or plot them using a computer.

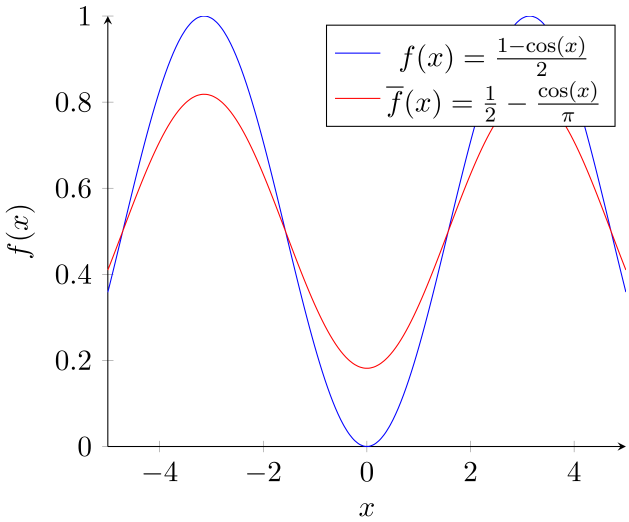
4.1.1 i

Calculate $\overline{f}(x)$. Show your work.

$$\begin{aligned}
\overline{f}(x) &= \frac{1}{\pi}\int_{x''=x-\frac{\pi}{2}}^{x''=x+\frac{\pi}{2}} f(x'')dx'' \\
&= \frac{1}{2}(1 - \frac{2}{\pi}\cos(x)) \\
&= \frac{1}{2} - \frac{cos(x)}{\pi}
\end{aligned}$$

4.1.2 ii

A plot shows that $f(x'')$ is a negative cosine function that extends from 0 to 1 in the vertical with a mean of 1/2. $\overline{f}(x'')$ is a smoothed version of $f(x'')$.



4.1.3 iii

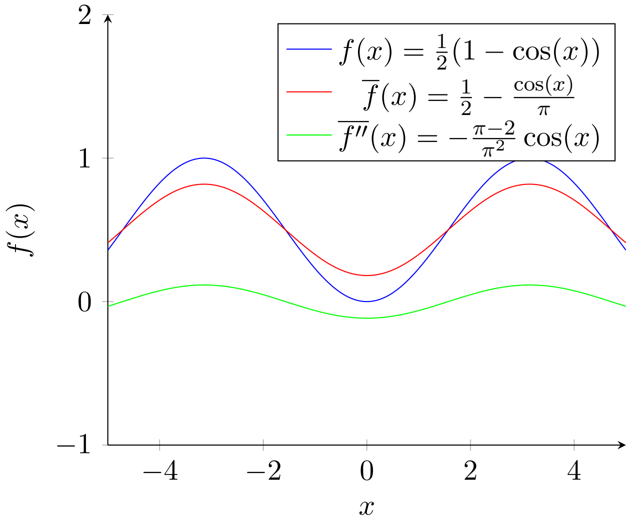
Suppose we define $f'' \equiv f(x'') - f(x)$. (Note that the definition here uses $f(x)$ rather than $f(x'')$.) Calculate $f''(x)$, showing your work.

$$\begin{aligned}
\overline{f''}(x) &= \frac{1}{\pi}\int_{x''=x-\frac{\pi}{2}}^{x''=x+\frac{\pi}{2}} f''(x'')dx'' \\
&= \frac{1}{\pi}\int_{x''=x-\frac{\pi}{2}}^{x''=x+\frac{\pi}{2}} (f(x'') - \overline{f}(x))dx'' \\
&= \overline{f}(x) - \overline{f}(x) \\
&= 0
\end{aligned}$$

4.1.4 iv

Now suppose we define $f'' \equiv f(x'') - f(x'')$. (Note that this time the definition uses $f(x'')$.) Calculate $f''(x'')$ using this new definition, showing your work. Plot $f''(x)$.

$$\begin{aligned}
\overline{f''}(x) &= \frac{1}{\pi}\int_{x''=x-\frac{\pi}{2}}^{x''=x+\frac{\pi}{2}} (f(x'') - \overline{f}(x''))dx'' \\
&= \overline{f}(x) - \overline{\overline{f}}(x) \\
\overline{\overline{f}}(x) &= \frac{1}{\pi}\int_{x''=x-\frac{\pi}{2}}^{x''=x+\frac{\pi}{2}} (\frac{1}{2} - \frac{\cos(x'')}{\pi})dx'' \\
&= \frac{1}{2}(1 - (\frac{2}{\pi})^2\cos(x)) \\
\overline{f''}(x) &= -\frac{\pi - 2}{\pi^2}\cos(x)
\end{aligned}$$



4.1.5 v

Explain in physical terms why $f''(x)$ from part iii) is different from $f''(x'')$ from part iv). Which part obeys Reynolds' rules?

Part iv) uses a running mean filter, and hence the average deviation from the running mean varies in space. In contrast, part iii) uses a constant (w.r.t. x'') filter, and hence the local deviations are larger, but they average to zero. Because the filter in part iii) is a constant, it obeys Reynolds' rules.

4.2 Problem Set 2

4.2.1 Problem 1: Cloud fraction for a "double delta function" PDF

Consider a volume of air that is half occupied by air with $r_t = r_{t1}$ and half occupied by air with $r_t = r_{t2}$. (Let $r_{t1} > r_{t2}$, and let $\Delta r_t \equiv r_{t1} - r_{t2}$.) Further suppose that the volume has uniform temperature everywhere and that the saturation mixing ratio is r_s. Assume that any vapor in excess of saturation is immediately converted to liquid, as we assumed earlier in the semester for non-precipitating, ice-free clouds.

(i) Write an expression for cloud fraction, C, in terms of r_{t1}, r_{t2}, r_s, and the Heaviside step function.

$$C \equiv \frac{1}{2}H(r_{t1} - r_s) + \frac{1}{2}H(r_{t2} - r_s)$$

(ii) Write an expression for liquid water mixing ratio, r_l, in terms of r_{t1}, r_{t2}, r_s, and the Heaviside step function.

$$r_l = \frac{1}{2}(r_{t1} - r_s)H(r_{t1} - r_s) + \frac{1}{2}(r_{t2} - r_s)H(r_{t2} - r_s)$$

4.2.2 Problem 2: Cloud fraction for a triangular PDF

Consider now a volume of air that has a symmetric PDF of $r_t, P(r_t)$, with the shape of an isosceles triangle. The PDF is piecewise linear function of r_t that becomes non-zero at $r_t = \overline{r_t} - \Delta r_t$, increases linearly to a peak at $r_t = \overline{r_t}$, and then decreases linearly to zero at $r_t = \overline{r_t} + \Delta r_t$. Assume that temperature is constant and that the saturation mixing ratio is r_s.

(i) Write the functional form of $P(r_t)$. Recall that the area under $P(r_t)$ must be 1.

$$P(r_t) = \begin{cases} 0 & , r_t \leq \overline{r_t} - \Delta r_t \\ \frac{r_t - \overline{r_t}}{\Delta r_t^2} + \frac{1}{\Delta r_t} & , \overline{r_t} - \Delta r_t \leq r_t < \overline{r_t} \\ -\frac{r_t - \overline{r_t}}{\Delta r_t^2} + \frac{1}{\Delta r_t} & , \overline{r_t} \leq r_t < \overline{r_t} + \Delta r_t \\ 0 & , r_t \geq \overline{r_t} + \Delta r_t \end{cases}$$

(ii) Write an expression for cloud fraction, C, in terms of $r_t, \Delta r_t$, and r_s.
In general,

$$C = \int P(r_t) H(r_t - r_s) dr_t$$

For the triangular PDF,

$$C = \begin{cases} 0 & , r_t \leq \overline{r_t} - \Delta r_t \\ \frac{1}{2}(\frac{r_s - (\overline{r_t} + \Delta r_t)}{\Delta r_t})^2 & , \overline{r_t} - \Delta r_t \leq r_t < \overline{r_t} \\ 1 - \frac{1}{2}(\frac{-r_s + (\overline{r_t} - \Delta r_t)}{\Delta r_t})^2 & , \overline{r_t} \leq r_t < \overline{r_t} + \Delta r_t \\ 0 & , r_t \geq \overline{r_t} + \Delta r_t \end{cases}$$

4.2.3 Problem 3: Sketch cloud fractions for the double delta and triangular PDFs

Overplot the two cloud fraction formulas that you have found as a function of r_t for fixed r_s and Δr_t. You may sketch by hand, if you wish.

Both formulas for C must asymptote to 0 for low values of r_t and asymptote to 1 for high values of r_t. However, the triangle PDF yields a smooth function of C, whereas the double delta formula yields discrete steps from $C = 0, 0.5, 1$. These steps are rather unrealistic, which is a drawback of the double delta PDF.

4.3 Problem Set 3

There is no Problem Set 3.

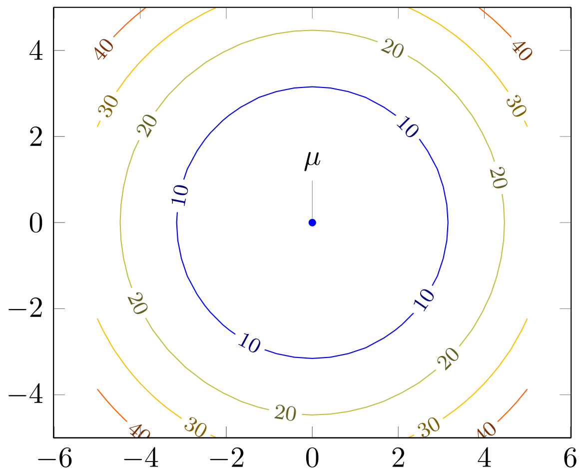
4.4 Problem Set 4

4.4.1 Problem 1: Conditional distribution of a 2D normal distribution

Consider a normal distribution with two variates, x_1 and x_2. Assume that the distribution has mean $\mu = 0$ and standard deviation $\sigma = 1$. Therefore, we can call the distribution $\mathcal{N}(x_1, x_2|0, 1)$.

(I) Assume that the covariance, Σ_{12}, between x_1 and x_2 is 0.

(a) Sketch contours of $\mathcal{N}(x_1, x_2|0, 1)$.

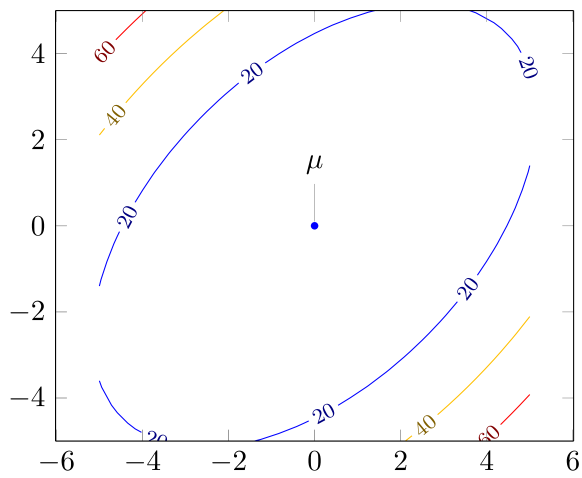


(b) What is the conditional average $E(X_1|X_2 = 1)$? Is $E(X_1|X_2 = 1) > \mu_1$, or is $E(X_1|X_2 = 1) < \mu_1$, or is $E(X_1|X_2 = 1) = \mu_1$?

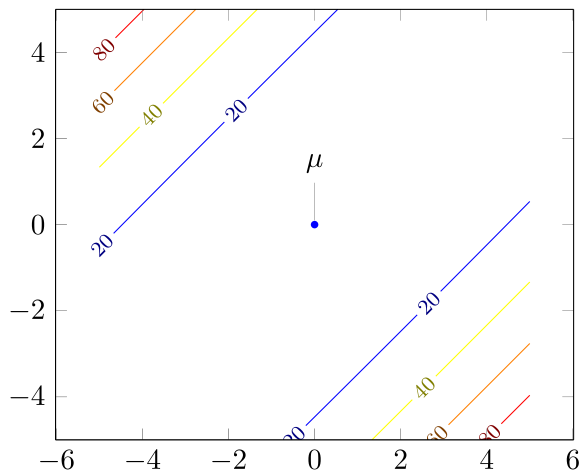
$$\mathbb{E}(X_1|X_2 = 1) = \mu_1, \text{ because the covariance is 0.}$$

(c) What is the standard deviation σ of the conditional distribution? Is it less than or greater than the standard deviation of $\mathcal{N}$, which is 1? Sketch the conditional distribution.
The standard deviation σ of the conditional distribution is unchanged at 1, again because there is zero covariance.

(II) Now repeat (a), (b), and (c), but assume that $\Sigma_{12} = 0.5$.
In this case, the covariance is positive. So the conditional mean is positive ($= 0 + 0.5(1-1)(1-0) = 0.5$) and the conditional variance is less than $1 (= 1 - 0.5(1-1)0.5)$.

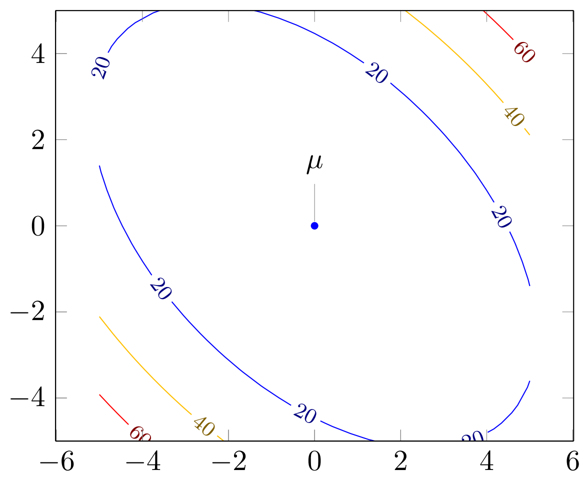


(III) Repeat (a), (b), and (c), but assume that $\Sigma_{12} = 0.99$.



(IV) The conditional average in (III) is larger than that in (II) because the correlation concentrates the values at high values. The standard deviation of the conditional distribution in (III) is less than that in (II) because the correlation concentrates the points along the diagonal.

(V) Repeat (a), (b), and (c), but assume that $\Sigma_{12} = -0.5$.
With negative covariance, the conditional mean is negative, but the conditional variance is still reduced.



4.5 Problem Set 5

4.5.1 Problem 1: Transformation of samples by use of the Cholesky decomposition

Suppose that a covariance matrix is given by

$$\Sigma = \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix}$$

Then the Cholesky matrix L is given by

$$L = \begin{pmatrix} 1 & 0 \\ r_{12} & \sqrt{1 - r_{12}^2} \end{pmatrix}$$

1. Verify, by direct computation, that $LL^T = \Sigma$.

$$\begin{pmatrix} 1 & 0 \\ r_{12} & \sqrt{1 - r_{12}^2} \end{pmatrix} \cdot \begin{pmatrix} 1 & r_{12} \\ 0 & \sqrt{1 - r_{12}^2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 \cdot 1 + 0 \cdot 0 & 1 \cdot r_{12} + 0 \cdot \sqrt{1 - r_{12}^2} \\ 1 \cdot r_{12} + 0 \cdot \sqrt{1 - r_{12}^2} & r_{12} \cdot r_{12} + \sqrt{1 - r_{12}^2} \cdot \sqrt{1 - r_{12}^2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & r_{12} \\ r_{12} & r_{12}^2 + (1 - r_{12}^2) \end{pmatrix}$$

$$= \begin{pmatrix} 1 & r_{12} \\ r_{12} & 1 \end{pmatrix} = \Sigma$$

2. Now suppose that the desired PDF is positively correlated, with $r_{12} = 0.99$, and it has zero mean: $\mu = 0$. Use the formula

$$X = L \cdot Y + \mu$$

to sketch how the Cholesky decomposition L maps the following four points, $Y = [1,1]^T, [1,1]^T, [-1,-1]^T, [-1,1]^T$ — into X. Does the distribution of points X look positively correlated? Does it matter whether X, Y, and μ are treated as row or column vectors?

3. Repeat part 2), but let $r_{12} = -0.99$. Does the resulting distribution look negatively correlated?

4. What is the Cholesky decomposition L for the covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_x^2 & r_{12}\sigma_x\sigma_y \\ r_{12}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

Answer:

$$LL^T = \Sigma$$

$$\begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \cdot \begin{pmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{pmatrix} \overset{!}{=} \begin{pmatrix} \sigma_x^2 & r_{12}\sigma_x\sigma_y \\ r_{12}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

$$\implies L = \begin{pmatrix} \sigma_x & 0 \\ r_{12}\sigma_y & \sigma_y\sqrt{1 - r_{12}^2} \end{pmatrix}$$

4.6 Problem Set 6

4.7 Problem Set 7

4.7.1 Problem 1: Optimal importance sampling density, q(x)

Suppose that our subgrid distribution is given by $P(x) = \mathcal{N}(0, \sigma^2)$ and our autoconversion parameterization is given by $f(x) = H(x)x^2$.

i) Find the optimal importance sampling density, $q(x)$. Be sure that $q(x)$ is normalized.

$$q(x) \propto |f(x)|P(x) = H(x)x^2 \exp(-\frac{x^2}{2\sigma^2})$$

The area under $|f(x)|P(x)$ is $\frac{1}{2}\sqrt{2\pi}\sigma^3$. (To see this, let $\alpha \equiv \frac{1}{2\sigma^2}$ and write $x^2 \exp(-\alpha x^2)$ as $-\frac{\delta}{\delta\alpha}(\exp(-\alpha x^2))$.) We must divide by this factor in order to normalize $q(x)$. Therefore,

$$q(x) = \frac{2}{\sqrt{2\pi}\sigma^3} H(x)x^2 \exp(-\frac{x^2}{2\sigma^2})$$

ii) Find the value of x where the optimal $q(x)$ reaches a maximum. Is it greater than or less than the mean, μ, for $P(x)$?
The maximum of $q(x)$ occurs where $\frac{\delta q(x)}{\delta x} = 0$. This is at $x = \sqrt{2}\sigma$. This is greater than $\mu = 0$, assuming that $\sigma > 0$.