Optimal blending of multiple independent prediction models

Preprint · May 2021

1 author:

Peter Taraba

Optimal blending of multiple independent prediction models

Peter Taraba

June 2, 2021

Abstract

We derive blending coefficients for the optimal blend of multiple independent prediction models.

1 Introduction

Participants of the Netflix competition used model blending quite heavily - see for example [1]. Ensemble modeling is used also in [2]. In this work, we derive blending coefficients based on variances of different models with only the assumption of model independence.

Let $\hat{y}_{k,j}$ be a prediction of model $k \in [1, N]$ for element $j \in [1, M]$, where N is the number of different independent models and M is the number of measurements we have:

$$\hat{y}_{k,j} = y_j + r_{k,j},$$

where y_j is an expected prediction and $r_{k,j}$ is a random variable with normal distribution $R_k \sim \mathcal{N}(0, \sigma_k^2)$, which has a zero average (expected value of variable is 0). In this work, we derive optimal blending coefficients α_k such that blended prediction $\hat{y}_B$ is optimal:

$$\hat{y}_{B,j} = \sum_{k=1}^{N} \alpha_k \hat{y}_{k,j} = y_j \sum_{k=1}^{N} \alpha_k + \sum_{k=1}^{N} \alpha_k r_{k,j} = y_j + \sum_{k=1}^{N} \alpha_k r_{k,j}$$

with minimum variance σ_B^2, where $\sum_{k=1}^{N} \alpha_k = 1$.

2 Blending two independent models

Here we present two independent models

$$\hat{y}_{1,j} = y_j + r_{1,j}$$

$$\hat{y}_{2,j} = y_j + r_{2,j},$$

where $R_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $R_2 \sim \mathcal{N}(0, \sigma_2^2)$. We derive $\hat{\alpha} \in [0, 1]$ for which we get the optimal blending model

$$\hat{y}_{B,j} = \alpha(y_j + r_{1,j}) + (1 - \alpha)(y_j + r_{2,j}) = y_j + \alpha r_{1,j} + (1 - \alpha) r_{2,j}.$$

It is well known fact that a random variable combining two random variables $\alpha R_1 + (1 - \alpha) R_2$, where $R_1 \sim \mathcal{N}(0, \sigma_1^2)$, $R_2 \sim \mathcal{N}(0, \sigma_2^2)$ and R_1 and R_2 are independent, has a normal distribution $\mathcal{N}(0, \sigma_B^2 = \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2)$. For the mean we get:

$$E(Y_B) = \frac{1}{M} \sum_{j=1}^{M} (\alpha r_{1,j} + (1 - \alpha) r_{2,j}) = \alpha E(R_1) + (1 - \alpha) E(R_2) = 0$$

and for the variance we get:

$$E(Y_B^2) = \frac{1}{M} \sum_{j=1}^{M} (\alpha r_{1,j} + (1 - \alpha) r_{2,j})^2 =$$

$$= \alpha^2 \frac{1}{M} \sum_{j=1}^{M} r_{1,j}^2 + 2\alpha(1 - \alpha) \frac{1}{M} \sum_{j=1}^{M} r_{1,j} r_{2,j} + (1 - \alpha)^2 \frac{1}{M} \sum_{j=1}^{M} r_{2,j}^2.$$

And finally as R_1 and R_2 are independent (covariance $\frac{1}{M} \sum_{j=1} r_{1,j} r_{2,j} = 0$ is zero) we can write:

$$\sigma_B^2 = E(Y_B^2) = \alpha^2 \sigma_1^2 + (1 - \alpha)^2 \sigma_2^2.$$

To find the optimal blending parameter we compute where a partial derivative of the new variance of the blended model is zero:

$$\frac{\partial \sigma_B^2}{\partial \alpha} = 2\hat{\alpha}\sigma_1^2 - 2(1 - \hat{\alpha})\sigma_2^2 = 0,$$

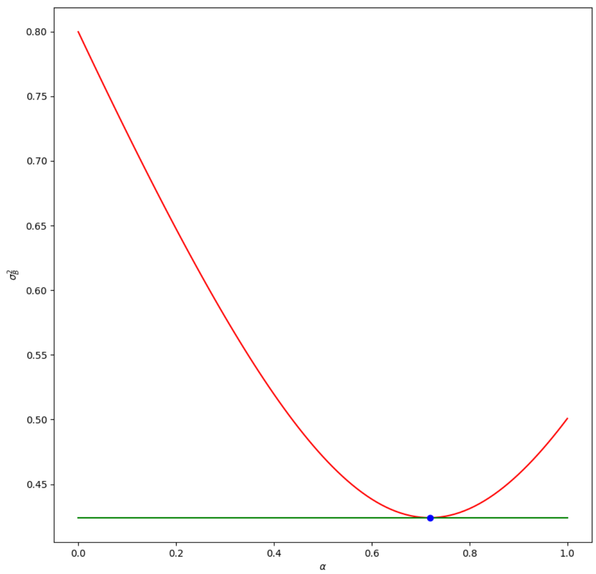

Figure 1: Red line - variance for different α. Green line - optimal variance σ_B^2. Blue dot - optimal α with its value $\sigma_B^2(\hat{\alpha})$.

from which

$$\hat{\alpha} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{1}$$

and the optimal variance will be:

$$\sigma_B^2(\hat{\alpha}) = \frac{\sigma_1^2\sigma_2^4}{(\sigma_1^2 + \sigma_2^2)^2} + \frac{\sigma_2^2\sigma_1^4}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}{(\sigma_1^2 + \sigma_2^2)^2} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{2}$$

In figure 1 we show how the variance is changing for different blending parameters α. Script is in appendix A or can be cloned from url: `https://github.com/pepe78/model-blending`. Blue dot - optimal value of blending parameter α matches simulation (minimal value for variance).

3 Blending three independent models

Now we consider three independent models

$$\hat{y}_{1,j} = y_j + r_{1,j}$$

$$\hat{y}_{2,j} = y_j + r_{2,j},$$

$$\hat{y}_{3,j} = y_j + r_{3,j},$$

where $R_1 \sim \mathcal{N}(0, \sigma_1^2)$, $R_2 \sim \mathcal{N}(0, \sigma_2^2)$ and $R_3 \sim \mathcal{N}(0, \sigma_3^2)$.

Here we blend optimally the first two models from the previous section:

$$\hat{y}_{4,j} = y_j + \hat{\alpha} r_{1,j} + (1 - \hat{\alpha}) r_{2,j} = y_j + r_{4,j},$$

where $R_4 \sim \mathcal{N}(0, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2})$ and then we find the blending parameter $\hat{\beta}$ for $\hat{y}_{3,j}$ and $\hat{y}_{4,j}$ such that

$$\hat{y}_{B,j} = y_j + \hat{\beta} r_{3,j} + (1 - \hat{\beta}) r_{4,j}. \tag{3}$$

Based on the equation 1 we get

$$\hat{\beta} = \frac{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}{\sigma_3^2 + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}.$$

Plugging this back into the equation 3 we get

$$\hat{y}_{B,j} = y_j + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} r_{3,j} + (1 - \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}) r_{4,j}$$

$$= y_j + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} r_{3,j} + \frac{(\sigma_1^2 + \sigma_2^2)\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} (\hat{\alpha} r_{1,j} + (1 - \hat{\alpha}) r_{2,j})$$

$$= y_j + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} r_{3,j} + \frac{(\sigma_1^2 + \sigma_2^2)\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} (\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} r_{1,j} + (1 - \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}) r_{2,j})$$

$$= y_j + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} r_{3,j} + \frac{\sigma_2^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} r_{1,j} + \frac{\sigma_1^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2} r_{2,j},$$

which is symmetrical, meaning model combination order is irrelevant. And finally for $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$ we get:

$$\hat{\alpha}_1 = \frac{\sigma_2^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}$$

$$\hat{\alpha}_2 = \frac{\sigma_1^2\sigma_3^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}$$

$$\hat{\alpha}_3 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_3^2 + \sigma_2^2\sigma_3^2}.$$

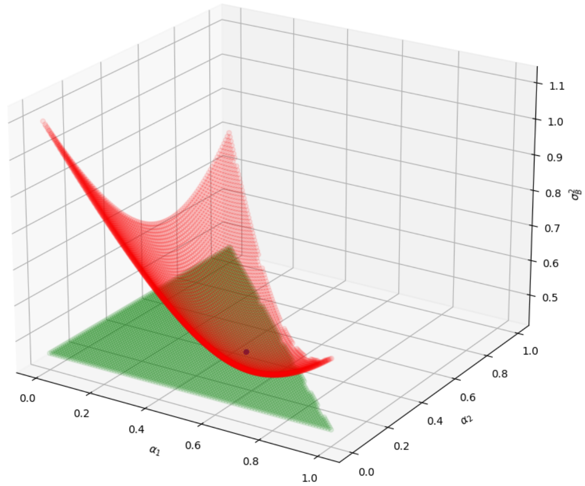

Figure 2: Red dots - variance for different α_1 and α_2. Green dots - optimal variance σ_B^2. Blue dot - optimal α_1 and α_2 and $1-\alpha_1-\alpha_2$ with its value $\sigma_B^2(\hat{\alpha})$.

Combining the second and third models first and then combining the result with the first model would lead to the same optimal blending parameters. The order of the combination is inconsequential. And for the final variance we get from equation 2

$$\sigma_B^2(\hat{\alpha}) = \frac{\sigma_3^2 \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}}{\sigma_3^2 + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}} = \frac{\sigma_1^2\sigma_2^2\sigma_3^2}{\sigma_1^2\sigma_2^2+\sigma_1^2\sigma_3^2+\sigma_2^2\sigma_3^2}.$$

In figure 2 we show how variance is changing for different blending parameters α_1 and α_2 and $\alpha_3 = 1-\alpha_1-\alpha_2$. Script is in appendix B. Blue dot - optimal value of blending parameters $(\alpha_1, \alpha_2, 1-\alpha_1-\alpha_2)$ matches simulation (minimal value for variance).

4 Blending N independent models

Now that we have formulas for two and three different models, we prove formulas for N independent models with normal distributions:

$$\hat{y}_{k,j} = y_j + r_{k,j},$$

where $R_k \sim \mathcal{N}(0, \sigma_k^2)$. We combine these models as follows:

$$\hat{y}_{B,j} = \sum_{k=1}^{N} \alpha_k \hat{y}_{k,j},$$

where $\sum_{k=1}^{N} \alpha_k = 1$ and $\alpha_k > 0$ for $k \in [1, N]$.

First we show model independence is still needed.

Lemma 1. *Having N independent models with normal distributions $R_k \sim \mathcal{N}(0, \sigma_k^2)$ for $k \in [1, N]$, when combined as $r_{B,j} = \sum_{k=1}^{N} \alpha_k r_{k,j}$, variance of R_B is $\sigma_B^2 = \sum_{k=1}^{N} \alpha_k^2 \sigma_k^2$.*

Proof.

$$\sigma_B^2 = E((\sum_{k=1}^{N} \alpha_k r_{k,j})^2) = \frac{1}{M}(\sum_{k=1}^{N}\sum_{j=1}^{M} \alpha_k^2 r_{k,j}^2 + 2\sum_{k=1}^{N}\sum_{\substack{l=1\\ l\neq k}}^{N}\sum_{j=1}^{M} \alpha_k \alpha_l r_{k,j} r_{l,j})$$

$$= \sum_{k=1}^{N} \alpha_k^2 \frac{1}{M}\sum_{j=1}^{M} r_{k,j}^2 + 2\sum_{k=1}^{N}\sum_{\substack{l=1\\ l\neq k}}^{N} \alpha_k \alpha_l \frac{1}{M}\sum_{j=1}^{M} r_{k,j} r_{l,j}$$

As models are independent (covariance $\frac{1}{M}\sum_{j=1}^{M} r_{k,j} r_{l,j} = 0$ is zero for $l \neq k$), we get

$$\sigma_B^2 = \sum_{k=1}^{N} \alpha_k^2 \frac{1}{M}\sum_{j=1}^{M} r_{k,j}^2 = \sum_{k=1}^{N} \alpha_k^2 \sigma_k^2,$$

which ends the proof. $\square$

Theorem 2. *Having N independent models with normal distributions $R_k \sim \mathcal{N}(0, \sigma_k^2)$ for $k \in [1, N]$, we get an optimal blend with parameters*

$$\hat{\alpha_k} = \frac{\prod_{\substack{j=1\\ j\neq k}}^{N} \sigma_j^2}{\sum_{i=1}^{N} \prod_{\substack{j=1\\ j\neq i}}^{N} \sigma_j^2},$$

and these independent models form normal distribution $\mathcal{N}(0, \sigma_B^2)$, which has variance

$$\sigma_B^2 = \frac{\prod_{j=1}^{N} \sigma_j^2}{\sum_{i=1}^{N} \prod_{\substack{j=1\\ j\neq i}}^{N} \sigma_j^2}.$$

Proof. For $N = 2$, we have shown it in section 2. Now we use induction, if it is true for N, then it is true also for $N + 1$.

Remark. *We have shown this also for three models in section 3, but as for induction it is not needed, section 3 is only a motivational section for how to derive final formulas for* N *models.*

We combine two normal distributions $\mathcal{N}(0, \frac{\prod_{j=1}^{N}\sigma_j^2}{\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2})$ (assuming it is true for N) and $\mathcal{N}(0, \sigma_{N+1}^2)$. From equation 2 (lemma 1 is incorporated in this equation) we get

$$\sigma_B^2 = \frac{\frac{\prod_{j=1}^{N}\sigma_j^2}{\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2}\sigma_{N+1}^2}{\frac{\prod_{j=1}^{N}\sigma_j^2}{\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2} + \sigma_{N+1}^2} = \frac{\sigma_{N+1}^2\prod_{j=1}^{N}\sigma_j^2}{\prod_{j=1}^{N}\sigma_j^2 + \sigma_{N+1}^2\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2}$$

$$= \frac{\prod_{j=1}^{N+1}\sigma_j^2}{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2}$$

and hence we have shown optimal variance is valid for $N + 1$. Now we must show the same for the optimal coeficients. From 1 we get

$$\hat{\alpha} = \frac{\sigma_{N+1}^2}{\frac{\prod_{j=1}^{N}\sigma_j^2}{\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2} + \sigma_{N+1}^2} = \frac{\sigma_{N+1}^2\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2}{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2}$$

and hence

$$\hat{\alpha}_{N+1} = 1 - \hat{\alpha} = 1 - \frac{\sigma_{N+1}^2\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2}{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2} = \frac{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2 - \sigma_{N+1}^2\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2}{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2}$$

$$= \frac{\prod_{\substack{j=1\\j\neq N+1}}^{N+1}\sigma_j^2 + \sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2 - \sigma_{N+1}^2\sum_{i=1}^{N}\prod_{\substack{j=1\\j\neq i}}^{N}\sigma_j^2}{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2} = \frac{\prod_{\substack{j=1\\j\neq N+1}}^{N+1}\sigma_j^2}{\sum_{i=1}^{N+1}\prod_{\substack{j=1\\j\neq i}}^{N+1}\sigma_j^2},$$

which proves $\hat{\alpha}_{N+1}$. And lastly to show the same for $\hat{\alpha_k}$ for $k \in [1, N]$:

$$\hat{\alpha_k} = \hat{\alpha}\frac{\prod_{\substack{j=1\\j\neq k}}^N \sigma_j^2}{\sum_{i=1}^N \prod_{\substack{j=1\\j\neq i}}^N \sigma_j^2} = \frac{\sigma_{N+1}^2 \sum_{i=1}^N \prod_{\substack{j=1\\j\neq i}}^N \sigma_j^2}{\sum_{i=1}^{N+1} \prod_{\substack{j=1\\j\neq i}}^{N+1} \sigma_j^2} \frac{\prod_{\substack{j=1\\j\neq k}}^N \sigma_j^2}{\sum_{i=1}^N \prod_{\substack{j=1\\j\neq i}}^N \sigma_j^2}$$

$$= \frac{\sigma_{N+1}^2}{\sum_{i=1}^{N+1} \prod_{\substack{j=1\\j\neq i}}^{N+1} \sigma_j^2} \prod_{\substack{j=1\\j\neq k}}^N \sigma_j^2 = \frac{\prod_{\substack{j=1\\j\neq k}}^{N+1} \sigma_j^2}{\sum_{i=1}^{N+1} \prod_{\substack{j=1\\j\neq i}}^{N+1} \sigma_j^2},$$

which ends the proof. □

5 Going to infinity

If we can generate infinite independent models with distributions $R_i \sim \mathcal{N}(0, \sigma^2)$ (same variance), the final variance will be

$$\sigma_B^2 = \lim_{N\to+\infty} \frac{\prod_{j=1}^N \sigma^2}{\sum_{i=1}^N \prod_{\substack{j=1\\j\neq i}}^N \sigma^2} = \lim_{N\to+\infty} \frac{\sigma^{2N}}{N\sigma^{2(N-1)}} = \lim_{N\to+\infty} \frac{\sigma^2}{N} = 0,$$

which means we can combine all these models to get a perfect prediction with no errors. Naturally, creating an infinite amount of independent models (with covariances zero) is a difficult if not impossible task in real applications.

Theorem 3. *Having N independent models with normal distributions $R_k \sim \mathcal{N}(0, \sigma_k^2)$ for $k \in [1, N]$ and their variances $\sigma_k^2 \leq \sigma_M^2$, where σ_M^2 is their maximum variance, combining them optimally with coefficients from the theorem 2, their combined variance is $\sigma_B^2 \leq \frac{\sigma_M^2}{N}$.*

Proof. We use induction again. For $N = 2$ we get

$$\sigma_B^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \leq \frac{\sigma_M^2}{2}$$

And this is true as

$$\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_2^2 \leq \sigma_M^2\sigma_1^2 + \sigma_M^2\sigma_2^2,$$

because

$$\sigma_1^2\sigma_2^2 \le \sigma_M^2\sigma_1^2$$

and

$$\sigma_1^2\sigma_2^2 \le \sigma_M^2\sigma_2^2.$$

Now if it is true for N, then it is true also for $N+1$. If

$$\sigma_{B,N}^2 \le \frac{\sigma_M^2}{N},$$

then

$$\sigma_{B,N+1}^2 \le \frac{\sigma_M^2}{N+1}.$$

And that is true as

$$\sigma_{B,N+1}^2 = \frac{\sigma_{B,N}^2\sigma_{N+1}^2}{\sigma_{B,N}^2 + \sigma_{N+1}^2} \le \frac{\sigma_M^2}{N+1},$$

because

$$N\sigma_{B,N}^2\sigma_{N+1}^2 + \sigma_{B,N}^2\sigma_{N+1}^2 \le \sigma_M^2(\sigma_{B,N}^2 + \sigma_{N+1}^2)$$

as both - this

$$N\sigma_{B,N}^2\sigma_{N+1}^2 \le \sigma_M^2\sigma_{N+1}^2$$

and this

$$\sigma_{B,N}^2\sigma_{N+1}^2 \le \sigma_M^2\sigma_{B,N}^2$$

are true, which ends the proof. □

This proof means, that if we combine infinite independent models with distributions $R_i \sim \mathcal{N}(0, \sigma_i^2)$, where variance $\sigma_i^2 \le \sigma_M^2$, we get variance:

$$\sigma_B^2 = \lim_{N\to+\infty} \frac{\prod_{j=1}^N \sigma_j^2}{\sum_{i=1}^N \prod_{\substack{j=1\\ j\ne i}}^N \sigma_j^2} \le \lim_{N\to+\infty} \frac{\sigma_M^2}{N} = 0.$$

Combining infinite independent models with bounded variances from above leads to perfect prediction with variance zero.

References

[1] Andreas Töscher, Michael Jahrer, and Robert M. Bell. The bigchaos solution to the netflix grand prize, 2009.

[2] Nina Schuhen, Thordis L. Thorarinsdottir, and Tilmann Gneiting. Ensemble model output statistics for wind vectors. *Monthly Weather Review*, 140(10):3204 – 3219, 2012.

A Appendix A

Script for figure 1:

```
import numpy as np
import matplotlib.pyplot as plt
import math

N = 100000
y = np.random.rand(N)

s1 = 0.5
y1 = y + np.random.normal(0.0, s1,(N))

s2 = 0.8
y2 = y + np.random.normal(0.0, s2,(N))

alphaOpti1 = s2*s2 / (s1*s1+s2*s2)
alphaOpti2 = s1*s1 / (s1*s1+s2*s2)
sOpti = math.sqrt(s1*s1*s2*s2/(s1*s1+s2*s2))
print('Computed', alphaOpti1, alphaOpti2, sOpti)

a1 = []
a2 = []
s = []
sO = []
for alpha in range(101):
    alphaX = alpha / 100.0
    yy = alphaX * y1 + (1.0 - alphaX) * y2
```

```
    a1.append(alphaX)
    a2.append(1.0-alphaX)
    s.append(math.sqrt(np.sum((yy-y)**2)/(N-1.0)))
    sO.append(sOpti)

pos = np.argmin(s)
print('Simulated', a1[pos], a2[pos], s[pos])

plt.figure(figsize=(10,10))
plt.plot(a1,s,'r')
plt.plot(a1,sO,'g')
plt.plot([alphaOpti1],[sOpti],'bo')
plt.xlabel(r"$\alpha$")
plt.ylabel(r"$\sigma_B^2$")
plt.show()
```

B Appendix B

Script for figure 2:

```
import numpy as np
import matplotlib.pyplot as plt
import math

N = 100000
y = np.random.rand(N)

s1 = 0.65
y1 = y + np.random.normal(0.0, s1,(N))

s2 = 0.8
y2 = y + np.random.normal(0.0, s2,(N))

s3 = 1.1
y3 = y + np.random.normal(0.0, s3,(N))

alphaOpti1 = s2*s2*s3*s3 / (s1*s1*s2*s2+s1*s1*s3*s3+s3*s3*s2*s2)
```

```
alphaOpti2 = s1*s1*s3*s3 / (s1*s1*s2*s2+s1*s1*s3*s3+s3*s3*s2*s2)
alphaOpti3 = s1*s1*s2*s2 / (s1*s1*s2*s2+s1*s1*s3*s3+s3*s3*s2*s2)

sOpti = math.sqrt(s1*s1*s2*s2*s3*s3/
        (s1*s1*s2*s2+s1*s1*s3*s3+s3*s3*s2*s2))
print('Computed:', alphaOpti1, alphaOpti2, alphaOpti3, sOpti)

a1 = []
a2 = []
a3 = []
s = []
sO = []
for alpha1 in range(101):
    alpha1X = alpha1 / 100.0
    for alpha2 in range(101):
        alpha2X = alpha2 / 100.0

        alpha3X = 1.0 - alpha1X - alpha2X
        if alpha3X >= 0:
            yy = alpha1X * y1 + alpha2X * y2 + alpha3X * y3

            a1.append(alpha1X)
            a2.append(alpha2X)
            a3.append(alpha3X)
            s.append(math.sqrt(np.sum((yy-y)**2)/(N-1.0)))
            sO.append(sOpti)

fig = plt.figure(figsize=(10,10))
ax = fig.add_subplot(111, projection='3d')

pos = np.argmin(s)
print('Simulated:', a1[pos], a2[pos], a3[pos], s[pos])

ax.scatter(a1,a2,s,c='r',alpha=0.1)
ax.scatter(a1,a2,sO,c='g',alpha=0.1)
ax.scatter([alphaOpti1],[alphaOpti2],[sOpti],c='b')
ax.set_xlabel(r"$\alpha_1$")
ax.set_ylabel(r"$\alpha_2$")
```

```
ax.set_zlabel(r"$\sigma_B^2$")
plt.show()
```