An Introduction to Probability and Stochastic Processes for Ocean, Atmosphere, and Climate Dynamics

1: Basic Probability

Adam Monahan

monahana@uvic.ca

School of Earth and Ocean Sciences, University of Victoria

Introduction

- "Climate is what you expect, weather is what you get."
  - Robert A. Heinlein

$\Rightarrow$ "expectation" lies at heart of notion of climate

$\Rightarrow$ this is a fundamentally probabilistic perspective

- Probability is a natural way of looking at atmosphere, ocean, and climate dynamics, both in terms of
  - statistics (probability and data)
  - stochastics (the dynamics of probability)
- These lectures will provide an overview of probability theory & stochastic processes in the context of atmosphere, ocean, and climate dynamics

Probability: Basic concepts

- Probabilities are used to characterise processes with indeterminate outcomes - that is, that are **random**
- Any given random process X is associated with a set of possible basic outcomes

$$\Omega = \{x_1, x_2, ..., x_n\}$$

  (which doesn't have to be discrete set)
- Can define an "event" A as any subset of these basic outcomes, and assign to that a **probability** $P(A)$ of its occurrence, with the rules:

$$\begin{array}{ll} (i) & 0 \leq P(A) \leq 1 \\ (ii) & P(\Omega) = 1 \text{ (something's got to happen)} \\ (iii) & P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \end{array}$$

(additivity of mutually exclusive events)

Probability densities

- Our main focus will be case of basic outcome space of continuous variables $\Omega = \{x = \omega \in \mathsf{R}\}$
- Define "differential" probability of x falling in small volume dx

$$dP(x_0) = P(x_0 \leq x \leq x_0 + dx)$$

- If $P(x)$ smooth, can define **probability density function (pdf)**

$$P(x_0 \leq x \leq x_0 + dx) = p_x(x)\, dx$$

  for which $\int p_x(x)\, dx = 1$
- Can also define *cumulative distribution function*

$$F(x_0) = P(x \leq x_0) = \int_{-\infty}^{x_0} p_x(x)\, dx$$

Joint and marginal pdfs

- Probability of A and B together is **joint probability**

$$P(A \text{ and} B) = P(A, B)$$

- For $\mathbf{x} \in \mathsf{R}^N$, $p_x(\mathbf{x}) = p_x(x_1, x_2, ..., x_N)$ is the **joint pdf** of multiple scalar random variables
- Can find pdf over subspace of $\mathbf{x}$ (the **marginal pdf**) by integrating out other variables
- E.g. in R^2:

$$\begin{aligned} p_{x_1}(x_1) &= \int p_x(x_1, x_2)\, dx_2 \\ p_{x_2}(x_2) &= \int p_x(x_1, x_2)\, dx_1 \end{aligned}$$

Conditional probabilities

- The **conditional probability** $P(A|B)$ measures the probability of A given that B is known to be true
- Conditional probabilities are related to joint probabilities:

$$P(A, B) = P(A|B)P(B) = P(B|A)P(A)$$

- $P(A|B)$ is a measure of the dependence of A and B
- **Independence**: B irrelevant to A, so $P(A|B) = P(A)$ and

$$P(A, B) = P(A)P(B)$$

- For x, y independent, joint pdf factors as product of marginals:

$$p_{xy}(x, y) = p_x(x)p_y(y) \quad \text{(independence)}$$

Bayes' Theorem

- From properties of conditional distributions

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{\int p(x, y)dx} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

$\Rightarrow$

$$p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

- Provides formal model of "knowledge acquisition" of system variable x through "experimental outcome" y:
  - $p(x)$ initial uncertain knowledge of x (the *prior*)
  - $p(y|x)$ predicted outcome of experiment
  - $p(x|y)$ improved knowledge of x given outcome of experiment y (the *posterior*)

Expectation and moments

- The pdf $p_x(\mathbf{x})$ defines a **measure** of relative probabilities, leading to the definition of the **expectation** (average) of a function $\mathbf{f}(\mathbf{x})$:

$$\mathsf{E}\{\mathbf{f}\} = \int \mathbf{f}(\mathbf{x})p_x(\mathbf{x})\, d\mathbf{x}$$

- Can interpret marginal pdf as "expectation" of conditional:

$$p_{x_1}(x_1) = \int p_x(x_1, x_2)\, dx_2 = \int p(x_1|x_2)p_{x_2}(x_2)\, dx_2$$

- Expectation used to define **moments**, which are useful characterisations of a pdf
- For scalar random variables:

$$\begin{aligned} \text{mean}(x) &= \mathsf{E}\{x\} && \text{pdf centre} \\ \text{var}(x) = \text{std}^2(x) &= \mathsf{E}\{(x - \mathsf{E}\{x\})^2\} && \text{pdf spread} \end{aligned}$$

Expectation and moments

- Two common higher-order (reduced) moments:

  - **skewness**

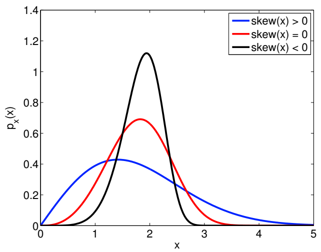
    $$\text{skew}(x) = \frac{\mathsf{E}\{(x - \mathsf{E}\{x\})^3\}}{\text{std}^3(x)}$$

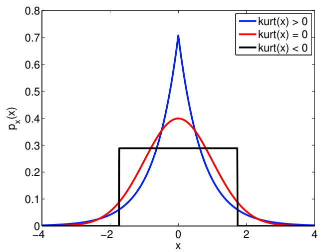
    measures *asymmetry* of pdf

  - **kurtosis**

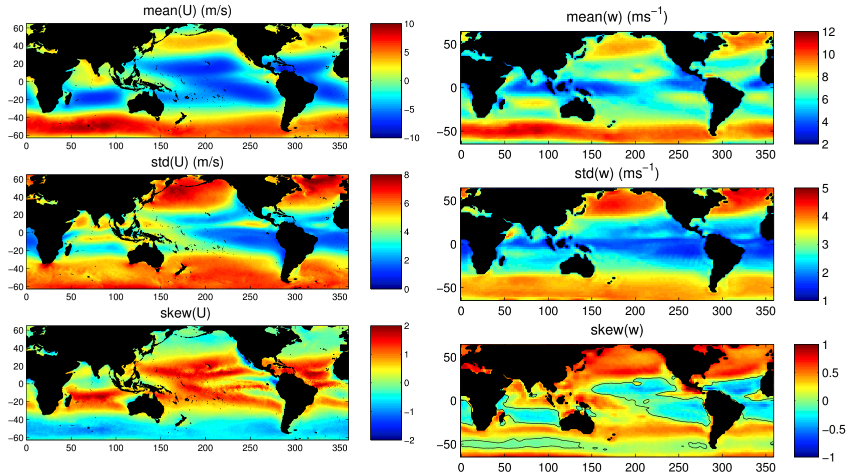
    $$\text{kurt}(x) = \frac{\mathsf{E}\{(x - \mathsf{E}\{x\})^4\}}{\text{std}^4(x)} - 3$$

    measures *flatness* of pdf





Moments of sea surface winds



Zonal Wind

Wind speed

Covariance and correlation

- Given two variables x and y with joint pdf $p_{xy}(x, y)$ define **covariance**

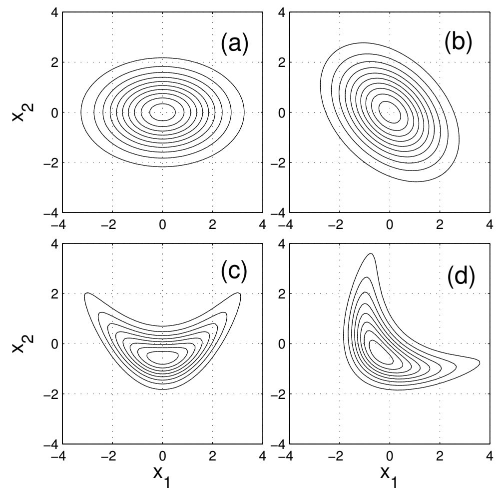
  $$\text{cov}(x, y) = \mathsf{E}\{(x - \mathsf{E}\{x\})(y - \mathsf{E}\{y\})\}$$

  and **correlation**

  $$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\text{std}(x)\text{std}(y)}$$

- These are measures of dependence between x and y
- If x and y are independent, covariance and correlation vanish - but **not vice versa!** (in general)

Covariance and correlation



(a) indep, uncorr
(b) dep, corr
(c) dep, uncorr
(d) dep, corr

Covariance matrix & EOFs

- For $\mathbf{x} \in \mathsf{R}^N$ define **covariance matrix**

$$\Sigma_{xx} = \mathsf{E}\left\{(\mathbf{x} - \mathsf{E}\{\mathbf{x}\})(\mathbf{x} - \mathsf{E}\{\mathbf{x}\})^T\right\}$$

  so $\mathrm{var}(\mathbf{x}) = \mathrm{diag}(\Sigma_{xx})$

- Efficiently described in terms of eigenvectors $\mathbf{e}_k$ known as **empirical orthogonal functions (EOFs)**:

$$\Sigma_{xx}\mathbf{e}_k = \lambda_k \mathbf{e}_k$$

- EOFs $\Rightarrow$ orthogonal basis in R^N which efficiently partitions variance
- In EOF basis new (uncorrelated) variables given by projections $a_k = \mathbf{x} \cdot \mathbf{e_k}$ with diagonal covariance matrix

$$(\Sigma_{aa})_{ij} = \lambda_j \delta_{ij}$$

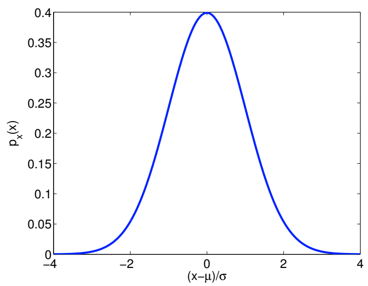
Families of pdfs

- There are many families of pdfs characterised by one or more parameters whose properties are well-studied
- The most important of these are the two-parameter Gaussian (or normal) distributions
- For a Gaussian scalar x

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

  such that

$$\mathrm{mean}(x) = \mu \quad \mathrm{std}(x) = \sigma$$
$$\mathrm{skew}(x) = 0 \quad \mathrm{kurt}(x) = 0$$



Multivariate Gaussian

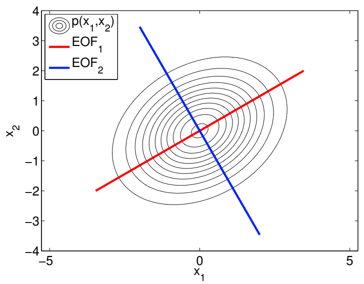
- Random vector $X \in \mathsf{R}^N$ is multivariate Gaussian when pdf is

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det \Sigma_{xx}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma_{xx}^{-1}(\mathbf{x}-\mu)\right)$$

  where

$$\mu = \mathrm{mean}(\mathbf{x})$$
$$\mathbf{\Sigma_{xx}} = \mathrm{cov}(\mathbf{x}, \mathbf{x})$$

- For a multivariate Gaussian
  - $\mathrm{corr}(x_i, x_j) = 0$ $\Rightarrow x_1, x_2$ independent
  - all marginals are Gaussian



Central limit theorem and law of large numbers

- One reason that Gaussian distributions are important is the **central limit theorem**, which states (loosely) that if x_n is a sequence of independent mean-zero random variables with finite variance, then

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i$$

  becomes Gaussian with finite variance as $n \to \infty$

- Another useful asymptotic result is the **law of large numbers**: for x_n a sequence of independent and identically distributed random variables

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_n \to \mathrm{mean}(x)$$

  That is: as n becomes very large, fluctuations in sum vanish

Maximum entropy pdfs

- Get moments from pdf; but how can we go the other way?
- The **entropy**

$$H = -\int p_x(\mathbf{x}) \ln p_x(\mathbf{x}) d\mathbf{x}$$

  measures “information content” of a pdf
- Given first N moments, can find pdf which maximises H subject to the constraint it must have specified moments; takes form

$$p_x(\mathbf{x}) = \frac{1}{Z} \exp\left(L_i x^i\right)$$

  where L_i are Lagrange multipliers of constrained optimisation
- Provides “least biased” pdf with given information
- Only mean and variance given, maximum entropy pdf *Gaussian*

Change of variables

- Suppose $y = f(x)$ where f is invertible
- Variable y is just another way of talking about x; intuitively expect

$$P(y_1 < y < y_2) = P(x_1 < x < x_2)$$

  if $y_1 = f(x_1)$ and $y_2 = f(x_2) \Rightarrow$ “conservation of probability”
- In terms of densities:

$$p_x(x)\ dx = p_y(y)\ dy = p_y(y)\ \frac{df}{dx}\ dx$$

  so

$$p_y(y) = \frac{p_x(f^{-1}(y))}{df/dx}$$

- For $\mathbf{x}, \mathbf{y} \in \mathrm{R}^N$

$$p_y(\mathbf{y}) = \frac{p_x(\mathbf{f}^{-1}(\mathbf{y}))}{\det(\partial f_i/\partial x_j)}$$

Change of variables: example

- Know the joint pdf of zonal and meridional wind components: $p_{uv}(u, v)$, but want marginal pdf of wind speed $p_w(w)$

1. move from (u, v) to polar coordinates (w, θ): $(u, v) = (w\cos\theta, w\sin\theta)$
2.

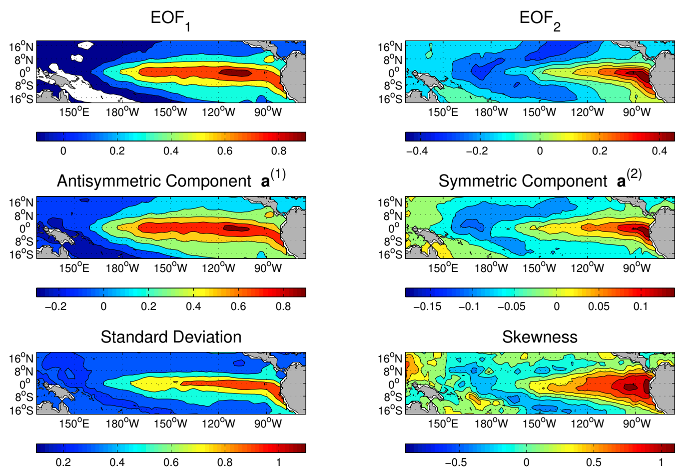
$$\begin{vmatrix} \partial_w u & \partial_\theta u \\ \partial_w v & \partial_\theta v \end{vmatrix} = \begin{vmatrix} \cos\theta & -w\sin\theta \\ \sin\theta & w\cos\theta \end{vmatrix} = w$$

$\Rightarrow$ $p_{w\theta}(w, \theta) = w p_{uv}(w\cos\theta, w\sin\theta)$

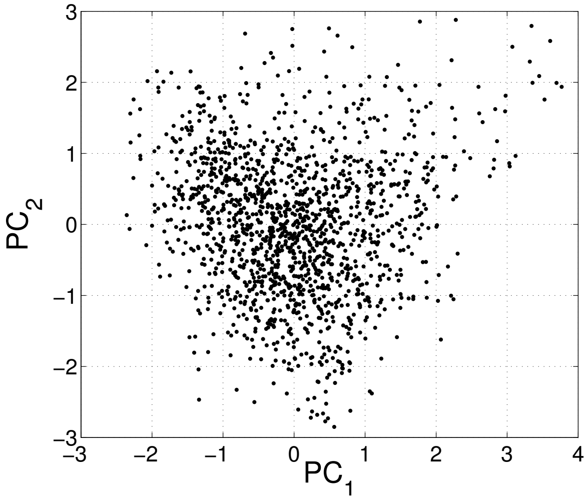
3. Integrate over θ to get marginal:

$$p_w(w) = w \int_{-\pi}^{\pi} p_{uv}(w\cos\theta, w\sin\theta)\ d\theta$$

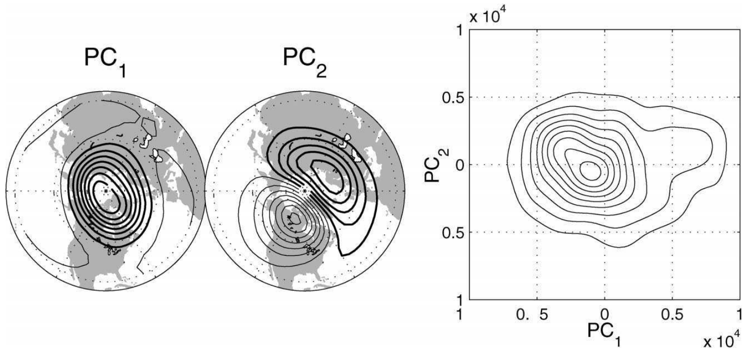
Case study: Tropical Pacific SST



Case study: Tropical Pacific SST



Case Study: 20 hPa geopotential height



References

- Gardiner, C.W. **Handbook of Stochastic Methods**, Springer
- Monahan, A.H. & A. Dai, 2004. “The spatial and temporal structure of ENSO nonlinearity”. *J. Clim*, **17**, 3026-3036.
- Monahan, A.H., J.C. Fyfe, & L. Pandolfo, 2003. “The vertical structure of wintertime climate regimes of the Northern Hemisphere extratropical atmosphere”. *J. Clim*, **16**, 2005-2021.
- von Storch, H. & F.W. Zwiers, **Statistical Analysis in Climate Research**, Cambridge University Press
- Wilks, D.S. **Statistical Methods in the Atmospheric Sciences**, Academic Press