SIO223A, Lecture 3, 01/14/2020

Probability and Random Variables

Cathy Constable
cconstable@ucsd.edu

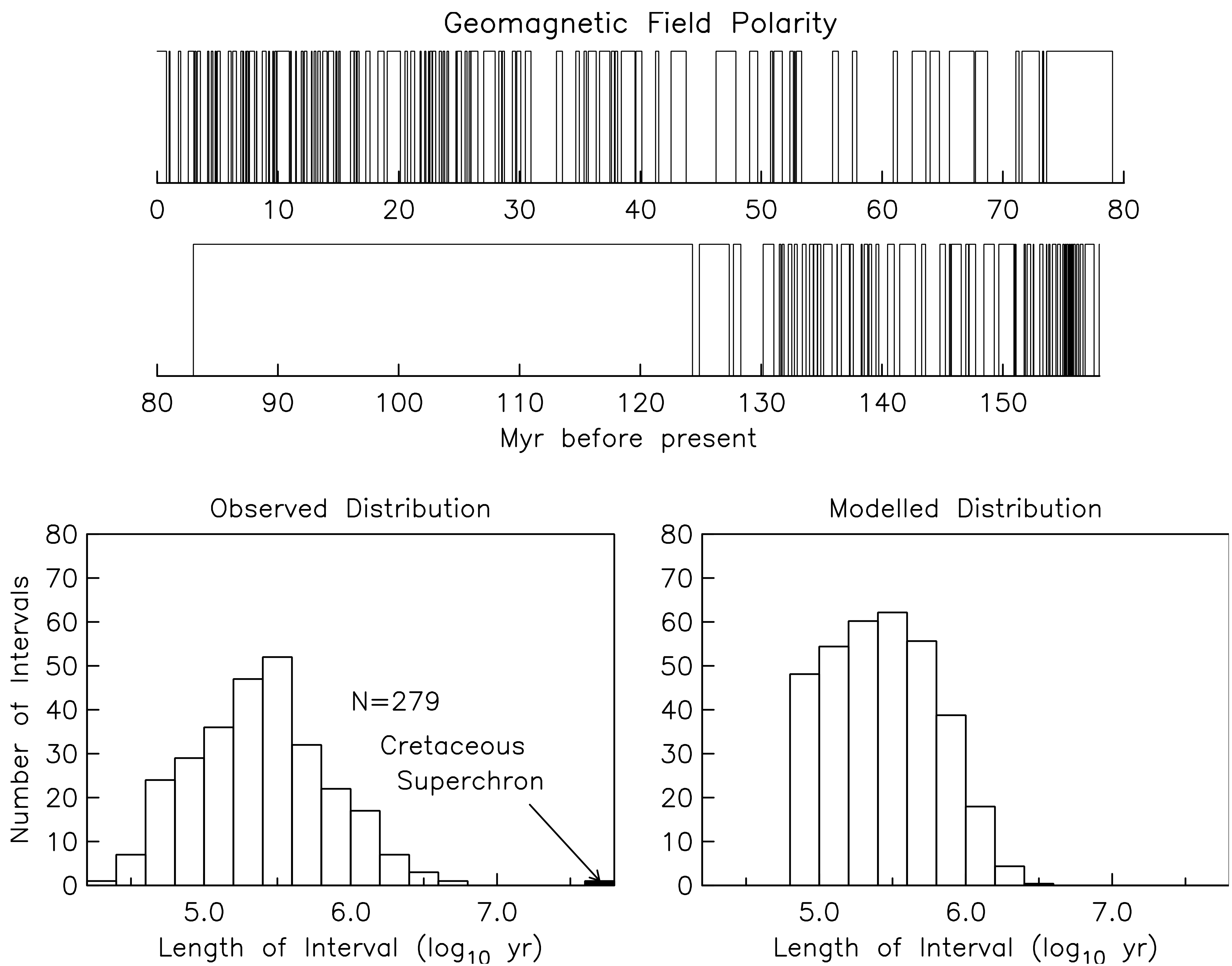
http://igppweb.ucsd.edu/~cathy/Classes/SIO223A/

Lecture 2 - recap

- Why are we here? Need for statistics
- Geophysical Examples -
  I. magnetic reversals, II. earthquakes
- Distances and error bounds
- Predicting Earthquakes

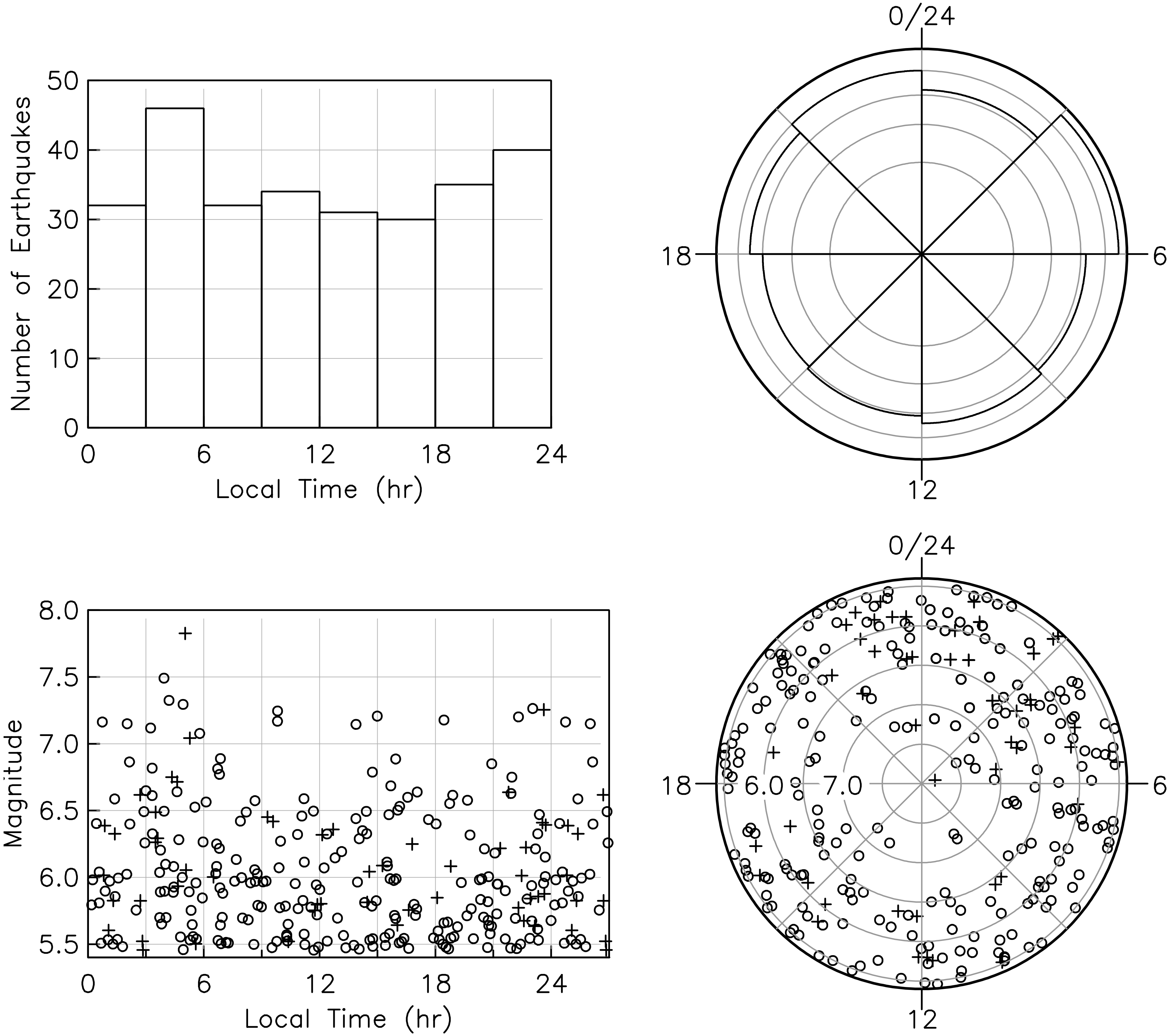
Terminology- Chapter 1

- histogram
- probability model
- stochastic model
- random variable
- probability theory
- statistics
- mean
- standard deviation
- estimates
- estimation theory
- point estimation
- robust estimation
- hypothesis test
- null hypothesis
- point process
- Poisson process
- Coxcomb plot
- confidence interval

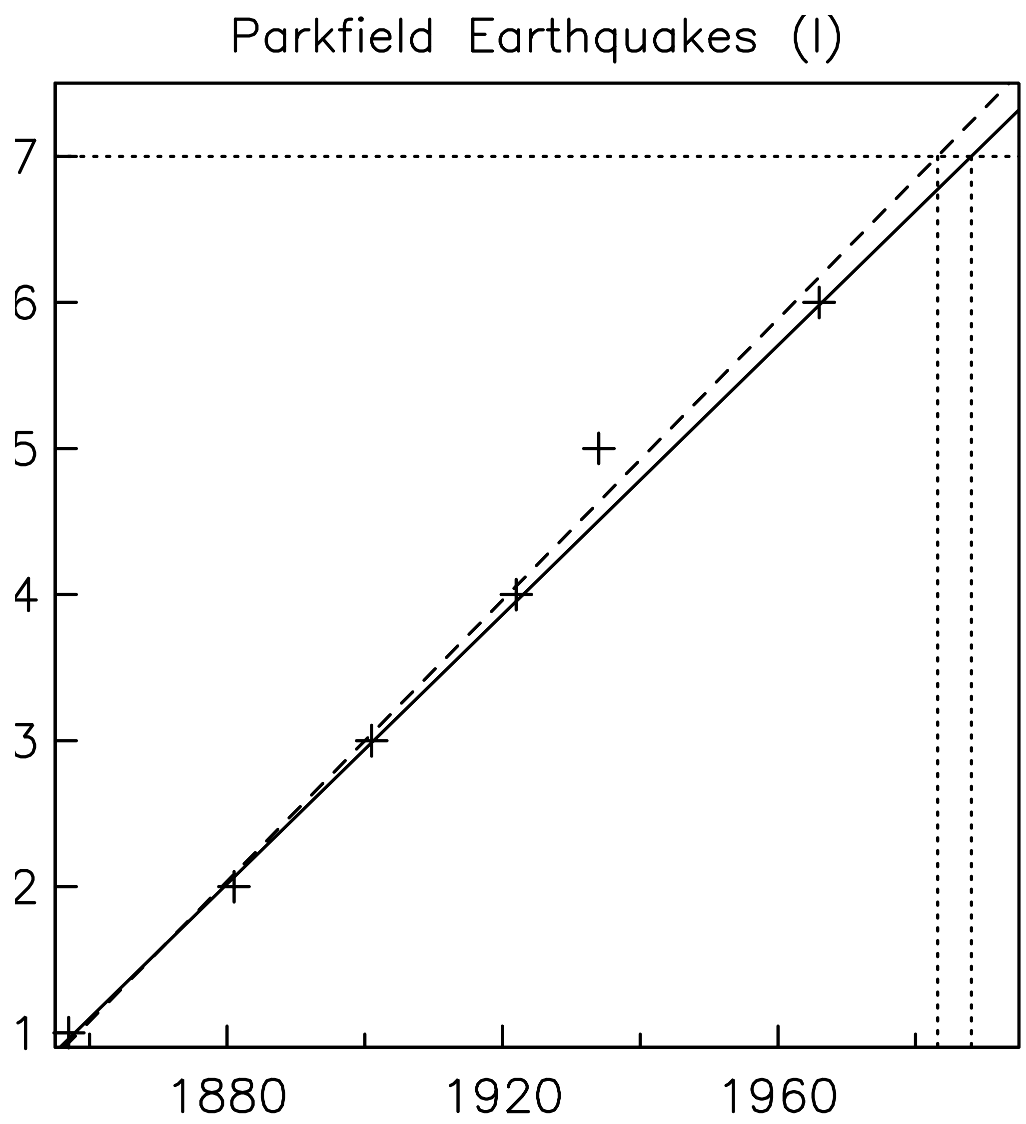

Geomagnetic Field Polarity
0
10
20
30
40
50
60
70
80
80
90
100
110
120
130
140
150
Myr before present
Observed Distribution
Number of Intervals
0
10
20
30
40
50
60
70
80
N=279
Cretaceous
Superchron
5.0
6.0
7.0
Length of Interval ($\log_{10}$ yr)
Modelled Distribution
0
10
20
30
40
50
60
70
80
5.0
6.0
7.0
Length of Interval ($\log_{10}$ yr)

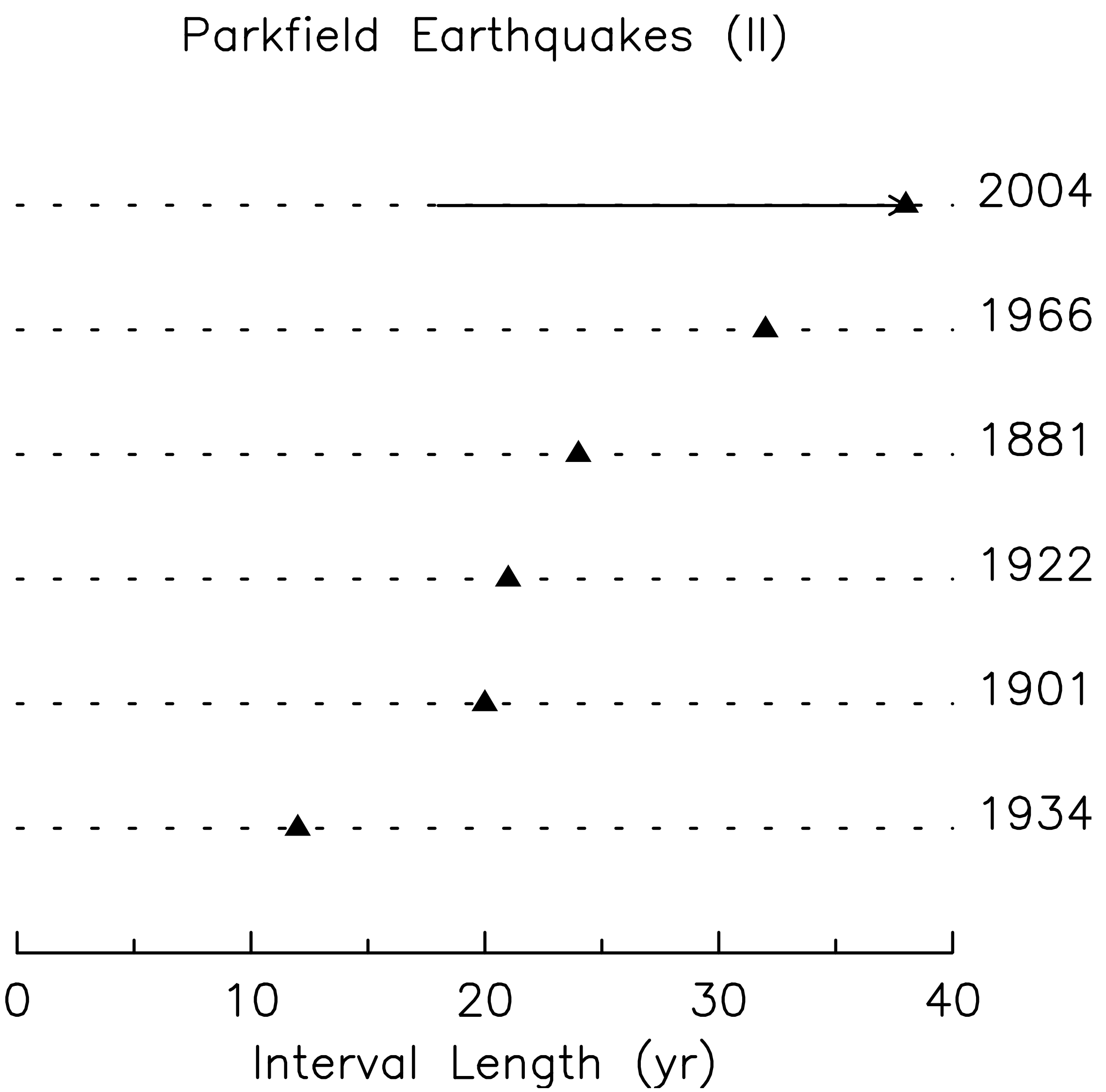
Times of California/Nevada Earthquakes

(1890–2012, magnitude 5.5 and above: 280 quakes)



Predicting Earthquakes?





Terminology- Chapter 2

- frequentist
- Bayesian
- subjective
- sample space
- probability axioms
- conditional probability
- independent
- Bayes' Theorem
- hypothesis
- prior probability
- likelihood
- posterior probability
- Bayesian inference
- random variable
- probability density function
- cumulative distribution function
- quantile

Chapter 2 Topics

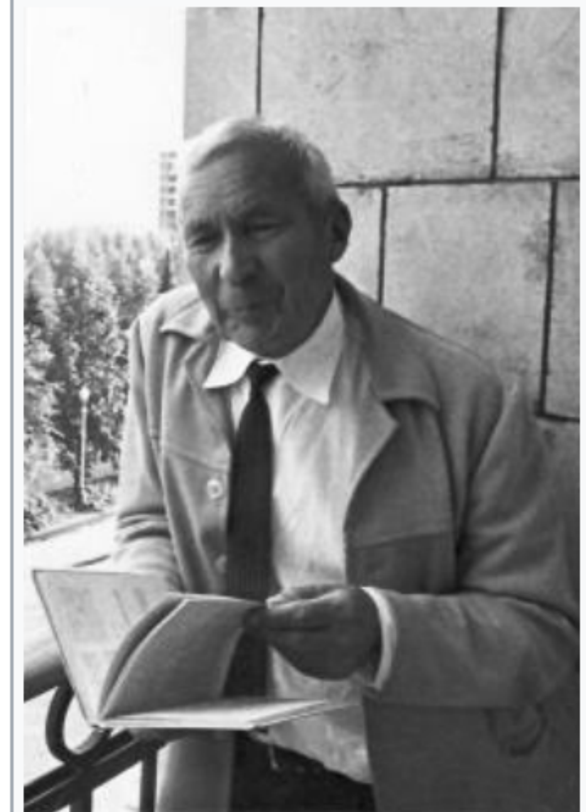
- What is probability? Frequentist vs Bayesian
- Basic axioms, sample space, and probability
- Conditional Probability
- Foreshock application
- Bayes' Theorem
- RVs: Probability Density and Distribution Functions

Basic Axioms of Probability

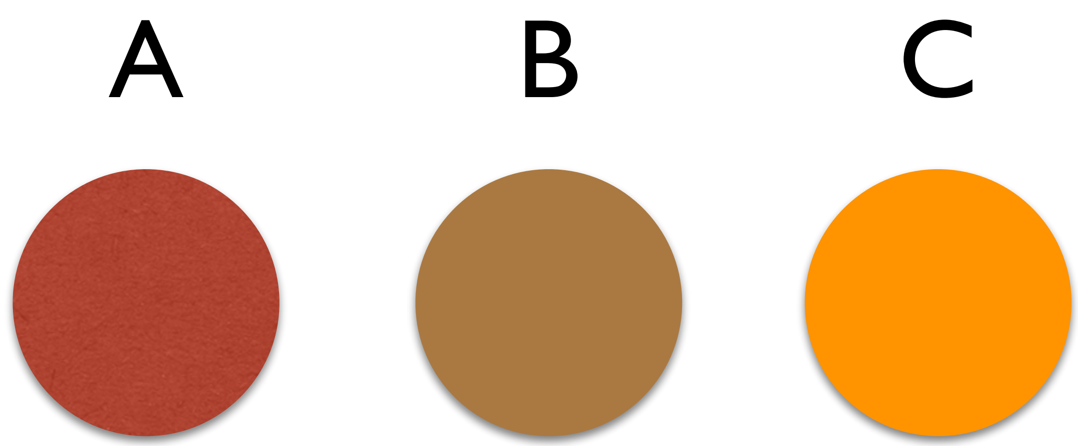
- https://en.wikipedia.org/wiki/Andrey_Kolmogorov

- Set theory: sample space, Ω, and outcome sets A, B, C, etc., with specified probability
- probability of all outcomes combined $\Pr(\Omega) = 1$
- probabilities are positive
- combinations of probabilities of disjoint outcomes can be summed

Andrey Kolmogorov



Born	Andrey Nikolaevich Kolmogorov 25 April 1903 Tambov, Russian Empire
Died	20 October 1987 (aged 84) Moscow, Soviet Union
Citizenship	Soviet Union
Alma mater	Moscow State University
Known for	Probability theory Topology Intuitionistic logic Turbulence studies Classical mechanics Mathematical analysis Kolmogorov complexity KAM theorem KPP equation
Spouse(s)	Anna Dmitrievna Egorova (m. 1942–1987)
Awards	Member of the Russian Academy of Sciences[1] Stalin Prize (1941) Balzan Prize (1962) ForMemRS (1964)[2] Lenin Prize (1965) Wolf Prize (1980) Lobachevsky Prize (1986)
Scientific career	
Fields	Mathematics
Institutions	Moscow State University
Doctoral advisor	Nikolai Luzin[3]



$$\Omega = A \cup B \cup C$$

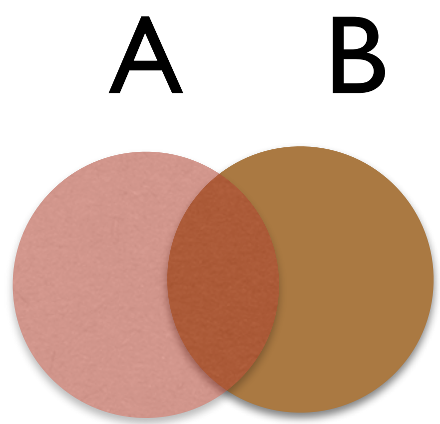
For disjoint sets A,B,C

$$Pr(\Omega) = Pr(A) + Pr(B) + Pr(C)$$

Andrey Kolmogorov

Born	Andrey Nikolaevich Kolmogorov 25 April 1903 Tambov, Russian Empire
Died	20 October 1987 (aged 84) Moscow, Soviet Union
Citizenship	Soviet Union
Alma mater	Moscow State University
Known for	Probability theory Topology Intuitionistic logic Turbulence studies Classical mechanics Mathematical analysis Kolmogorov complexity KAM theorem KPP equation
Spouse(s)	Anna Dmitrievna Egorova (m. 1942–1987)
Awards	Member of the Russian Academy of Sciences[1] Stalin Prize (1941) Balzan Prize (1962) ForMemRS (1964)[2] Lenin Prize (1965) Wolf Prize (1980) Lobachevsky Prize (1986)
	Scientific career
Fields	Mathematics
Institutions	Moscow State University
Doctoral advisor	Nikolai Luzin[3]

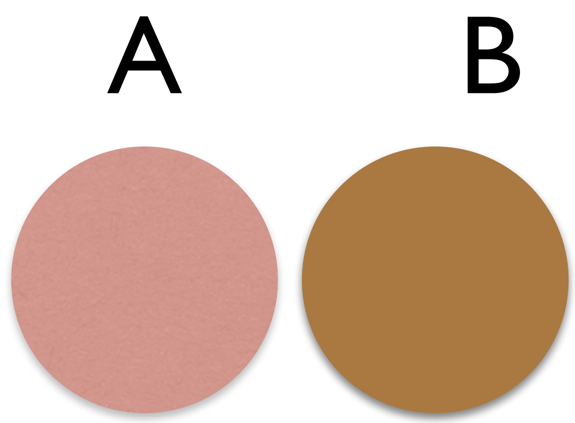
Conditional Probability



Outcomes A and B overlap

If B has already occurred,
what is the conditional prob of A?
We know that
$Pr(A|B)Pr(B) = Pr(A \cap B)$
so
$Pr(A|B) = Pr(A \cap B)/Pr(B)$

Independence



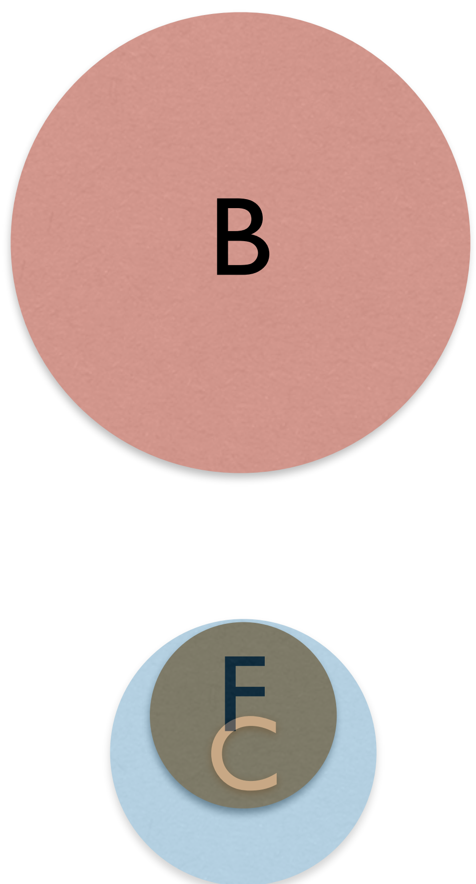
If $Pr(A|B) = Pr(A)$ then A and B are independent
and
$Pr(A \cap B) = Pr(A)Pr(B)$

SIO223A, Lecture 4, 01/16/2020

Today's Topics

- Bayes' Theorem
- RVs: Probability Density and Distribution Functions, Quantiles
- Empirical Cumulative Distributions
- Probability Plots and Q-Q Plots
- PDFs: mode, mean, variance, quartiles
- PDFs: moments, skewness, kurtosis,
- Functions of RVs: expectations, transformations, sums and products

Earthquake example



B is a background quake, which might occur frequently

F is a foreshock before a large quake

C is a large quake

Not all large quakes have a foreshock

What is prob of C given that one of either F or B has occurred?

Bayes' Theorem

Suppose we have N *disjoint* sets of outcomes, called $B_1, B_2, \dots B_N$, and another set A. The probability of both A and a particular one of the B's (say B_j) is, by the definition of conditional probability,

$$\Pr[A \cap B_j] = \Pr[B_j|A]\Pr[A] = \Pr[A|B_j]\Pr[B_j] \quad (2.7)$$

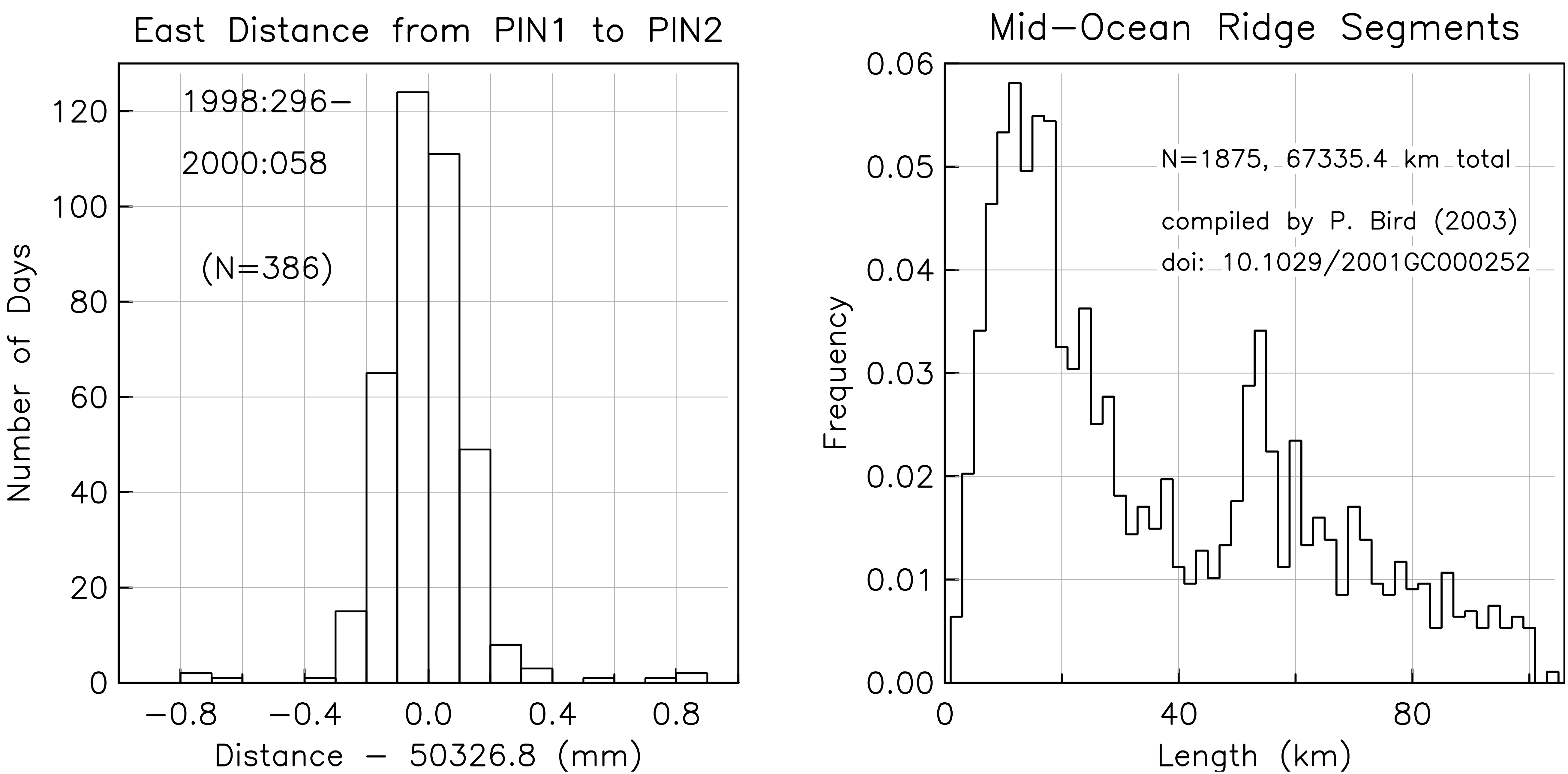
where you should remember that $\Pr[A \cap B_j] = \Pr[B_j \cap A]$. But, since the B's are disjoint, $\Pr[A] = \sum_j \Pr[A|B_j]\Pr[B_j]$. Combining this with 2.7, we find that

$$\Pr[B_j|A] = \frac{\Pr[A|B_j]\Pr[B_j]}{\sum_j \Pr[A|B_j]\Pr[B_j]} \quad (2.8)$$

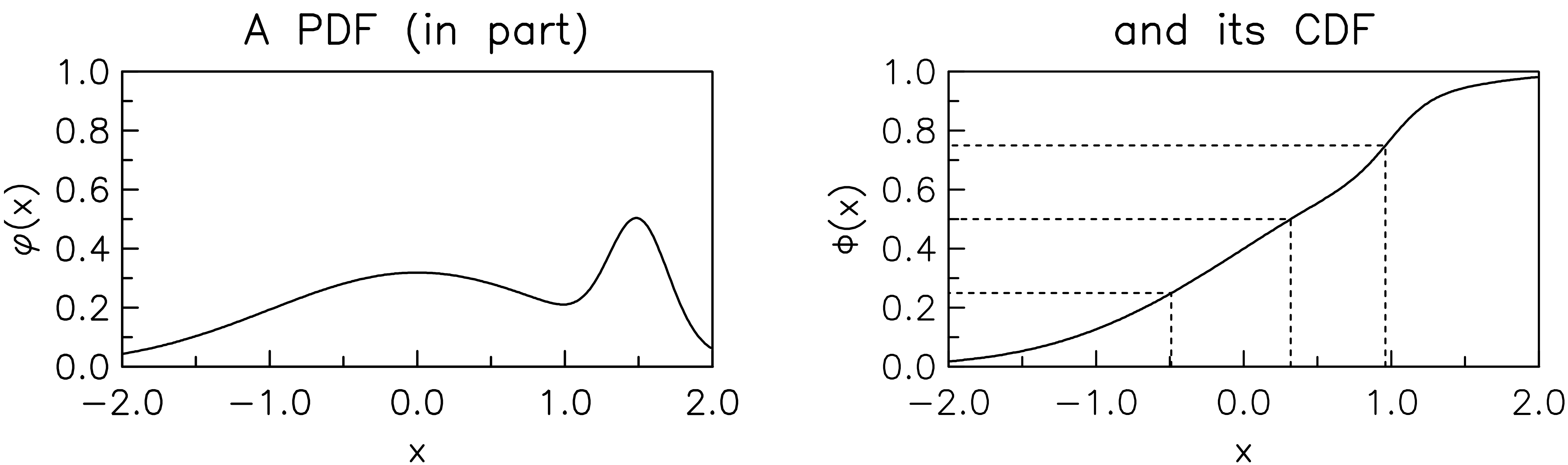
The different parts of this expression have special names: Each B_j is called a **hypothesis**, $\Pr[B_j]$ is called the **prior probability** of B_j, and $\Pr[A|B_j]$ the **likelihood** of A given B_j.

called the posterior probability for each hypothesis Bj

Histograms provide empirical estimates of probability



Continuous Random Variables X are specified by a probability density function

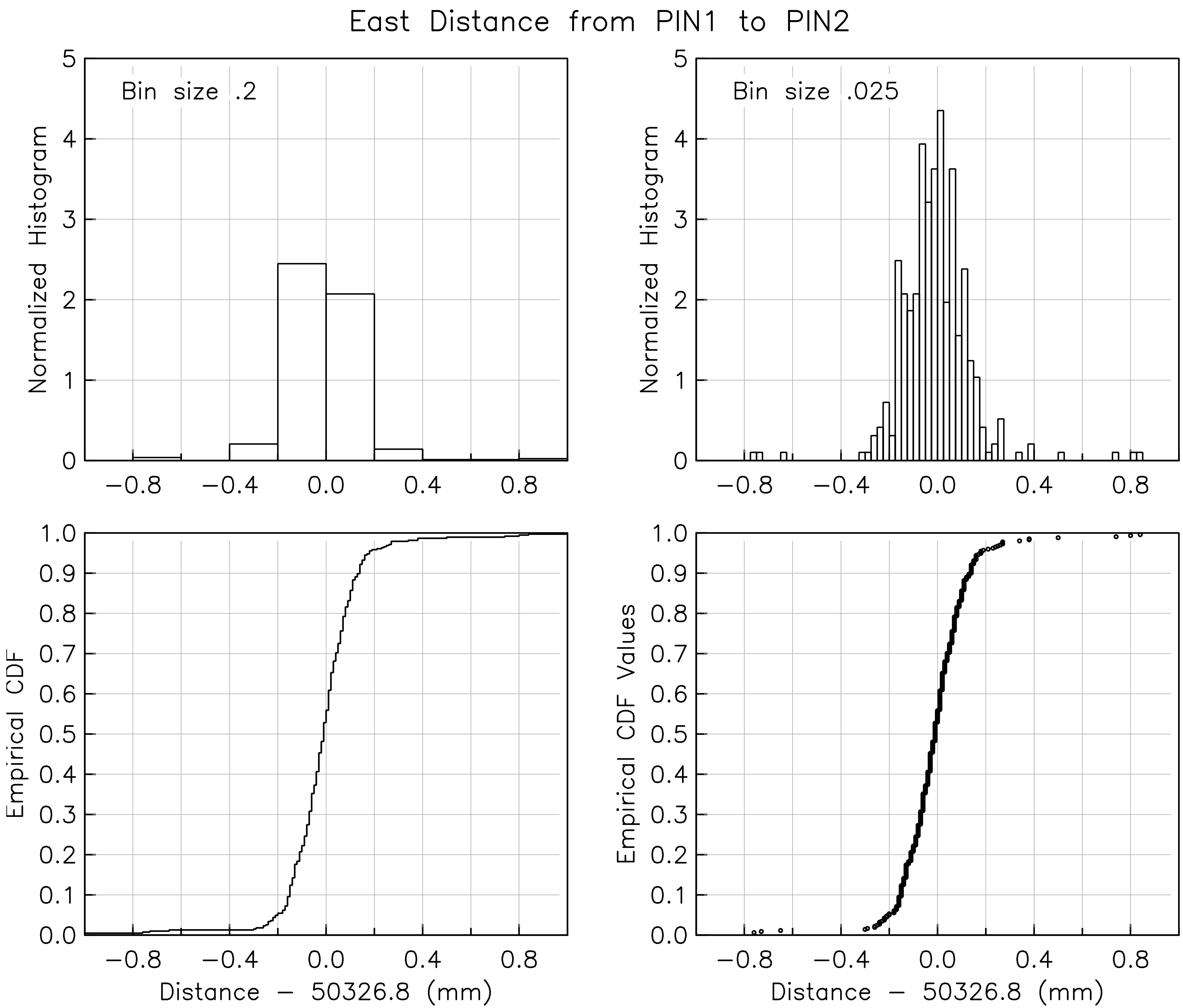


$$\Pr[x \le X \le x + \delta x] = \int_{x}^{x+\delta x} \phi(u)\,du$$

For any x and small interval δx this means

$$\Pr[x \le X \le x + \delta x] \approx \phi(x)\delta x + (\delta x)^2$$

so that $\phi(x)$ represents the density of probability per unit value of x in the neighborhood of x.


East Distance from PIN1 to PIN2
Bin size .2
Bin size .025
Normalized Histogram
Empirical CDF
Empirical CDF Values
Distance − 50326.8 (mm)
−0.8
−0.4
0.0
0.4
0.8

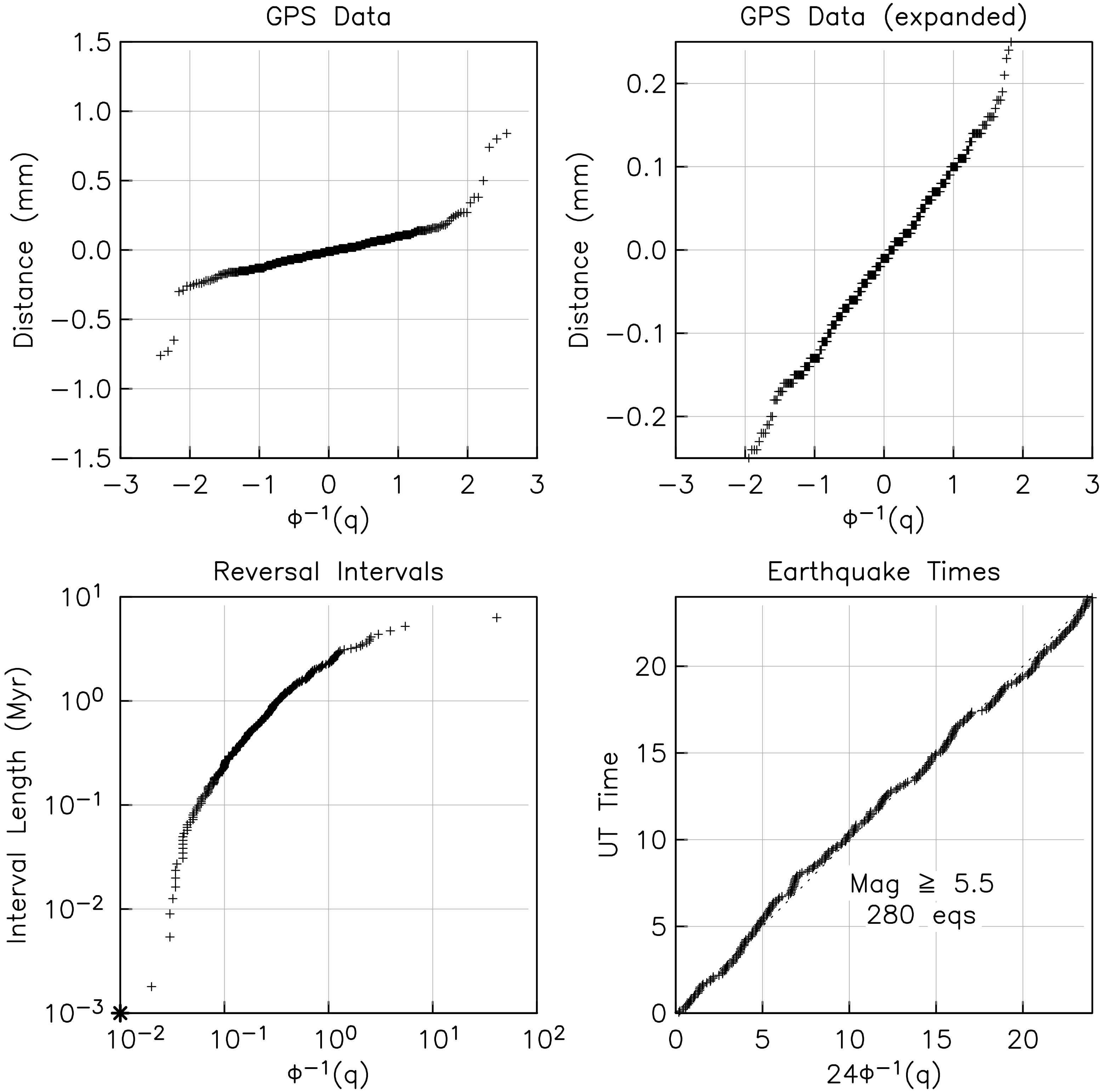
Probability density functions have the following properties:

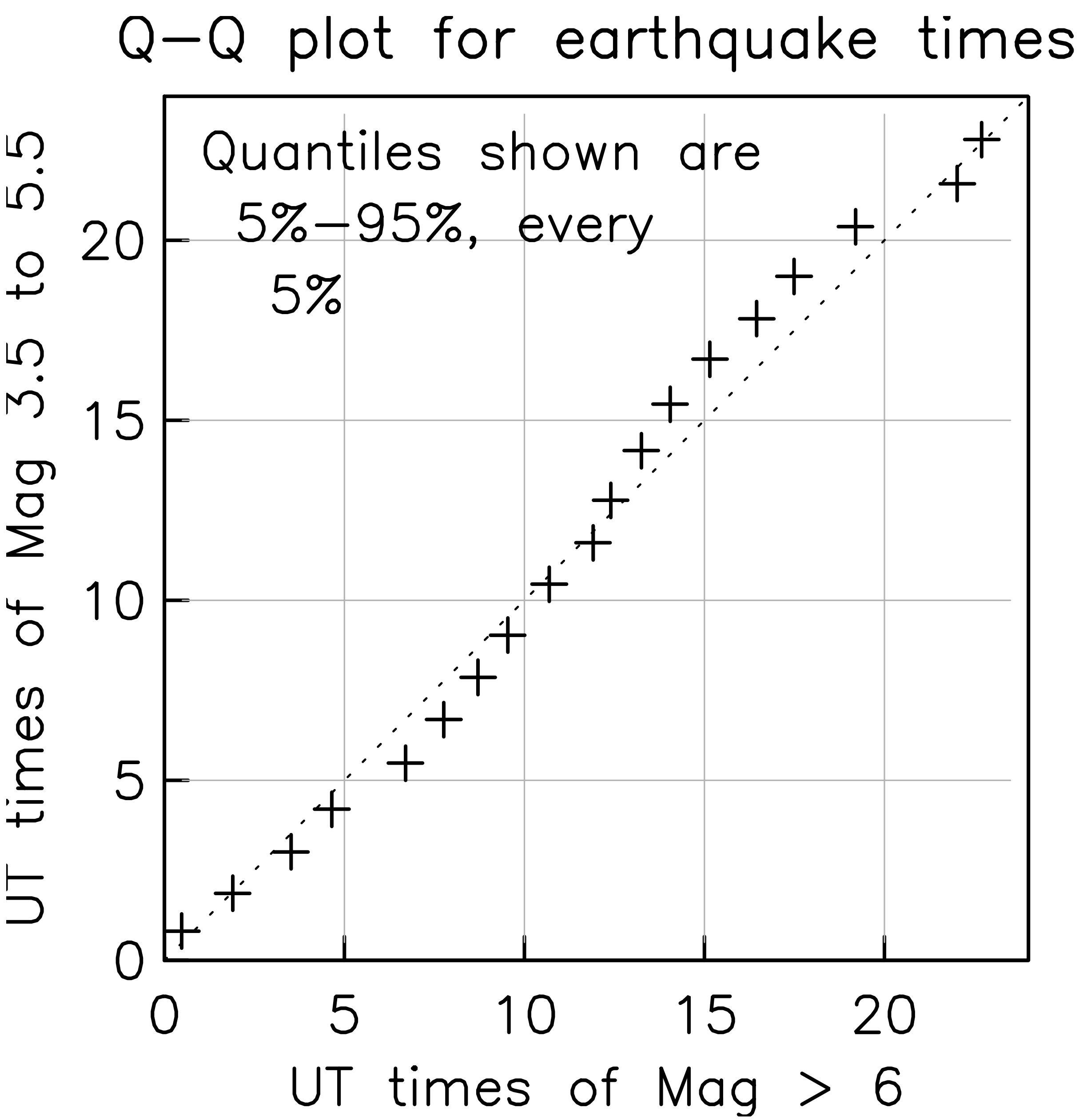
- $\phi(x) \geq 0$ for all x: probabilities are always positive.
- $\int_{L_b}^{L_t} \phi(x)\,dx = 1$: X must take on some value within its permissible range. Often this range is all of the real line, with $L_b = -\infty$ and $L_t = \infty$; but sometimes it is only a part. Section 1.2 already gave an example, which is that time intervals have to be positive, so $L_b = 0$ and $L_t = \infty$. Or, if we were considering the direction of something, X has to fall within $[0, 2\pi)$.

CDFs have the following properties

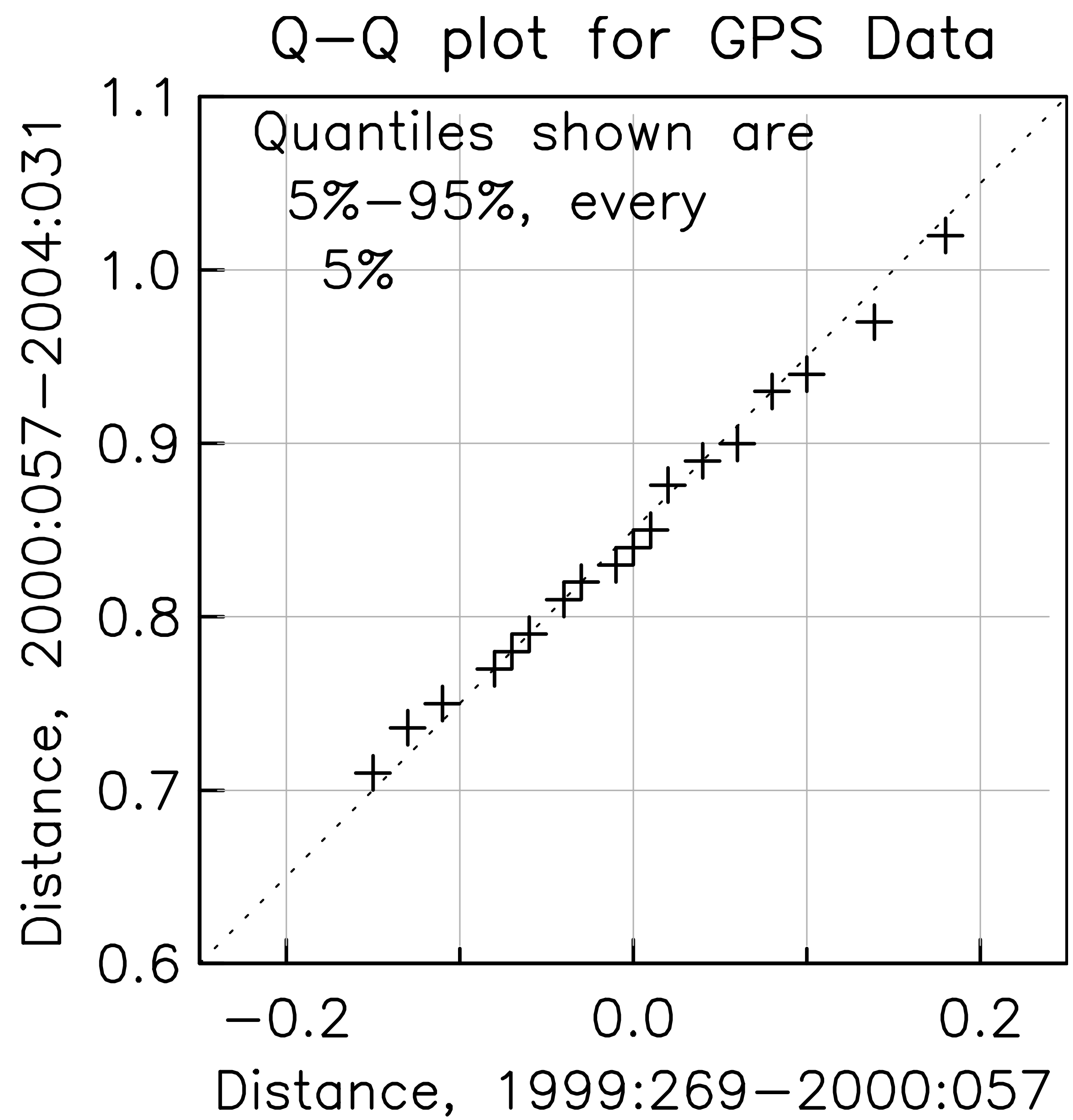
- $0 \leq \Phi(x) \leq 1$.
- $\lim_{x \to -\infty} \Phi(x) = 0 \quad \lim_{x \to \infty} \Phi(x) = 1$ or $\Phi(L_b) = 0$ and $\Phi(L_T) = 1$.
- Φ is non-decreasing; $\Phi(x+h) \geq \Phi(x) \qquad$ for $h \geq 0$.
- Φ is right continuous; $\lim_{h \to 0^+} \Phi(x+h) = \Phi(x)$; that is, as we approach any argument x from above, the function approaches its value at x.

Probability Plots

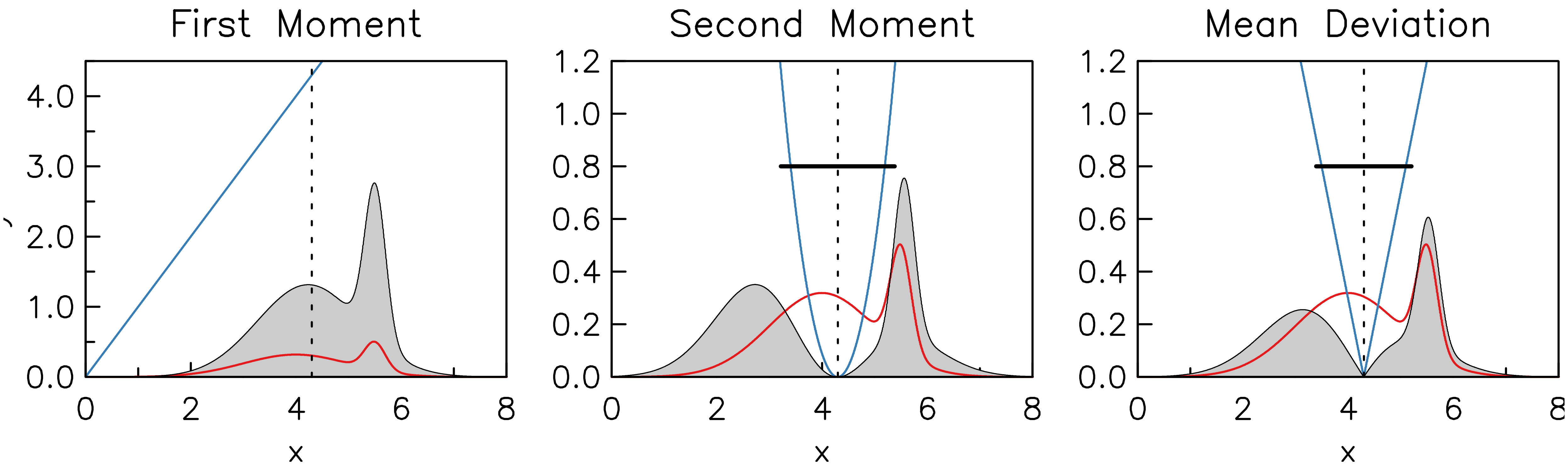




Q−Q plot for earthquake times
Quantiles shown are 5%−95%, every 5%
UT times of Mag 3.5 to 5.5
UT times of Mag > 6
0
5
10
15
20



Q−Q plot for GPS Data
Quantiles shown are 5%−95%, every 5%
Distance, 2000:057−2004:031
Distance, 1999:269−2000:057
0.6
0.7
0.8
0.9
1.0
1.1
−0.2
0.0
0.2


First Moment
Second Moment
Mean Deviation
0.0
1.0
2.0
3.0
4.0
0.2
0.4
0.6
0.8
1.2
0
2
4
6
8
x