Assignments

- Assignment: due Feb. 8
  - Read AMS draft policy statement on communicating science
  - http://www.ametsoc.org/policy/draftstatements/communicating_science_draftstatement.pdf
  - Summarize and critique the draft statement in a couple paragraphs
  - Comment in a couple paragraphs on the role of appropriate use of statistics for communicating science
  - Add what you would consider an appropriate summation of the use of statistics to communicate science in a couple of sentences

our possibilities are
ility of each event
uation and location.

ts

ty obvious, but
aticians involved,
oms", lemmas, etc.
any event is
n English: an event
or else it is not an

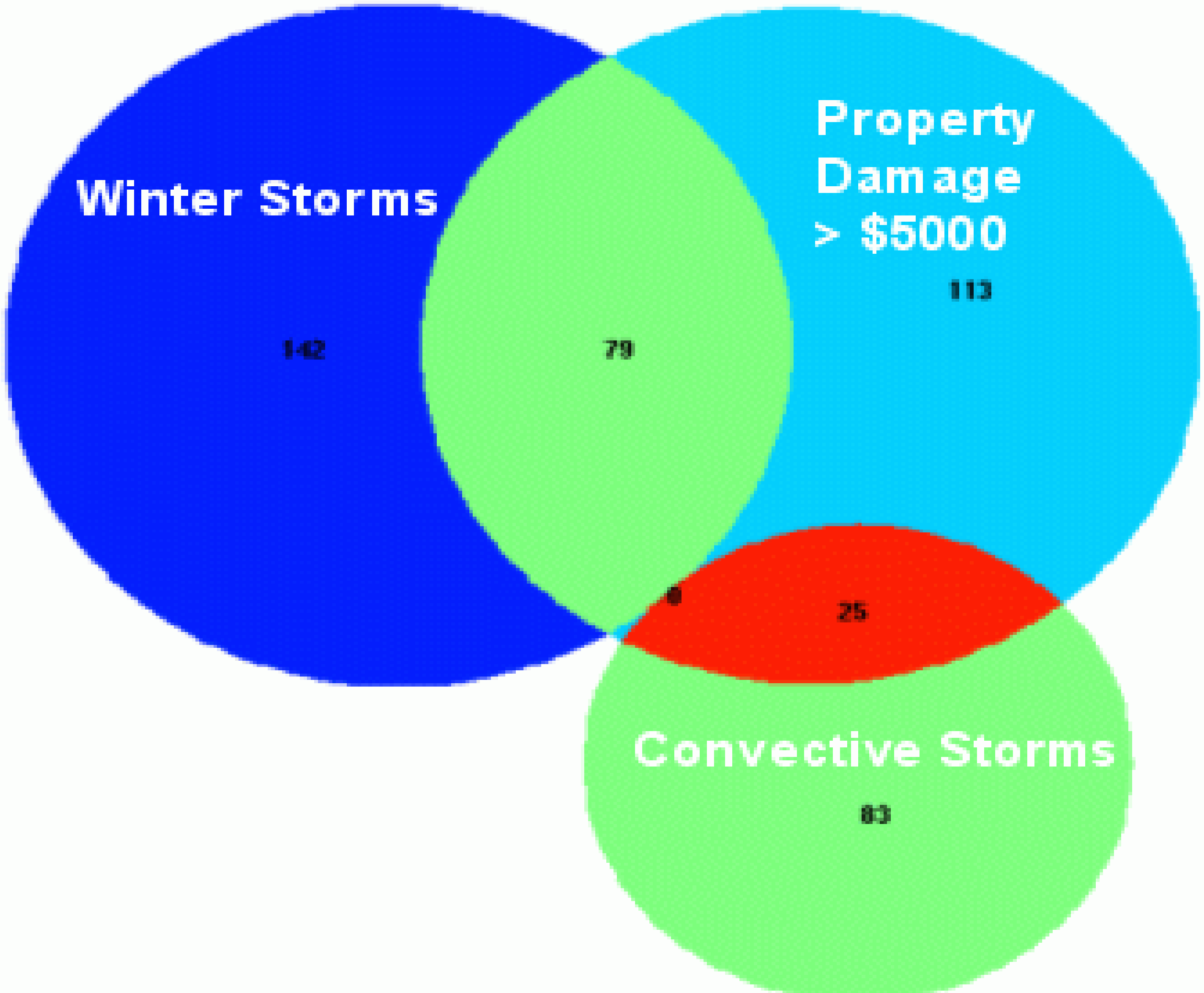
he compound event
. The probability
will happen is 1.

Temperature below Precipitation below	Temperature above Precipitation below
Temperature below Precipitation above	Temperature above Precipitation above

Figure 4.2. MECE possibilities for seasonal forecasts temperature and precpitation anomalies for a specific location.

t one or the other of two mutually exclusive events is the sum of their

Storm Reports: Salt Lake County 1993-2005



Number of Opportunities: 2340 (180 days * 13 years)

More concepts

- $\{E\}^c$- complement of $\{E\}$, event does not occur
- $Pr\{E\}^c = 1-\{E\}$
- $Pr\{E_1 \cap E_2\}$- joint probability that E_1 & E_2 occur
- $Pr\{E_1 \cap E_2\} = 0$ if E_1 & E_2 are mutually exclusive
- $Pr\{E_1 \cup E_2\}$- probability that E_1 OR E_2 occur
- $Pr\{E_1 \cup E_2\} = Pr\{E_1\} + Pr\{E_2\} - Pr\{E_1 \cap E_2\}$

Conditional Probability

- Conditional probability: probability of $\{E_2\}$ given that $\{E_1\}$ has occurred
- $Pr\{E_2 \mid E_1\} = Pr\{E_1 \cap E_2\} / Pr\{E_1\}$
- E_1 is the conditioning event
- If E_1 and E_2 are independent of each other, then

  $Pr\{E_2 \mid E_1\} = Pr\{E_2\}$ and $Pr\{E_1 \mid E_2\} = Pr\{E_1\}$
- Fair coin- Pr{heads} = 0.5
  - chance of getting heads on second toss is independent of the first

    Pr{heads | heads} = 0.5

    Pr{heads} twice = 0.5 * 0.5 = 025

Bayes Theorem

- $Pr\{E_2 \mid E_1\} = Pr\{E_1 \cap E_2\} / Pr\{E_1\}$ or
- $Pr\{E_1 \cap E_2\} = Pr\{E_2 \mid E_1\} * Pr\{E_1\}$
- $Pr\{E_1 \cap E_2\} = Pr\{E_1 \mid E_2\} * Pr\{E_2\}$ then
- $Pr\{E_1 \mid E_2\} = Pr\{E_2 \mid E_1\} * Pr\{E_1\} / Pr\{E_2\}$
- What is the advantage? Probability of conditioning event E_2 only computed once

Bayesian Application: how a rational person responds to evidence

	Pos Test	Neg Test	TOTAL
DRUG USER	0.495%	0.005%	0.5%
NOT DRUG USER	.995%	98.505%	99.5%
TOTAL	1.49%	98.51%	1

What are odds of falsely accusing non drug user?

E_1 – not drug user

E_2 - positive test

$Pr\{E_1\}$ – 99.5%

$Pr\{E_2\}$ – 1.49%

$Pr\{E_2 \mid E_1\}$ – .995%

$Pr\{E_1 \mid E_2\} = Pr\{E_2 \mid E_1\} * Pr\{E_1\} / Pr\{E_2\} =$

$= 0.995 * 99.5 / 1.49 = 66\%$

Conclusion: always ask for second opinion if clean and test positive

Conclusion: if you are a drug user, you only have a 33% chance of getting caught

Bayesian Application:

	COLD	WARM	TOTAL
DRY	20	60	80
WET	20	0	20
TOTAL	40	60	100

E_1 – cold
E_2 - dry
$Pr\{E_1\}$ – 0.4
$Pr\{E_2\}$ – 0.8
$Pr\{E_2 | E_1\}$ – 0.5

We can't tell if it is cold or warm
But we know it is dry
$Pr\{E_1 | E_2\} = Pr\{E_2 | E_1\} * Pr\{E_1\} / Pr\{E_2\} =$
$= 0.5 * 0.4 / 0.8 = 0.25$

NAME________________________

ATMOS 5040/6040 In Class Assignment

Number of opportunities: 2340

$\{E_1\}$ = occurrence of winter storms

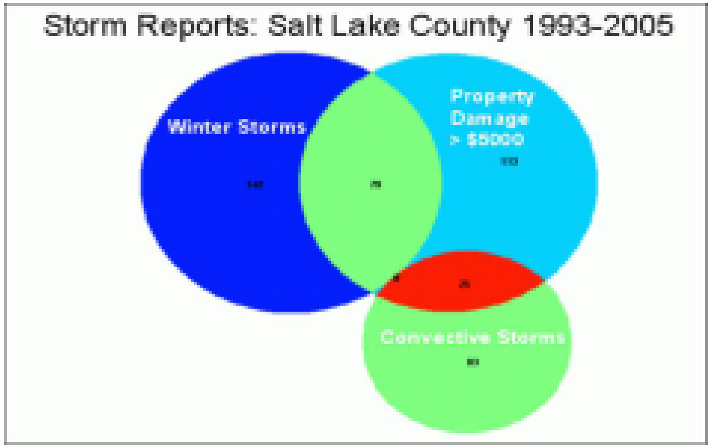
(142+79=221)

$\{E_2\}$ = occurrence of convective storms

(25+83=108)

$\{E_3\}$ = occurrence of property damage

(79+113+25= 217)



$Pr\{E_1\}$ = $Pr\{E_2\}$ = $Pr\{E_3\}$ = $Pr\{E_1 \cap E_2\}$ = $Pr\{E_2 \cap E_3\}$ = $Pr\{E_1 \cap E_3\}$=

$Pr\{E_1 \cup E_2\}$ = $Pr\{E_2 \cup E_3\}$ = $Pr\{E_1 \cup E_3\}$= $Pr\{E_3 \mid E_1\}$ = $Pr\{E_3 \mid E_2\}$ =

For a standard deck of 52 cards, deal out all cards in pairs. Pr{ace} = Pr{10-K} = Pr{2-9} =

What is the probability of getting a blackjack for any pair of cards: Pr{ ace ∩ 10-K} =

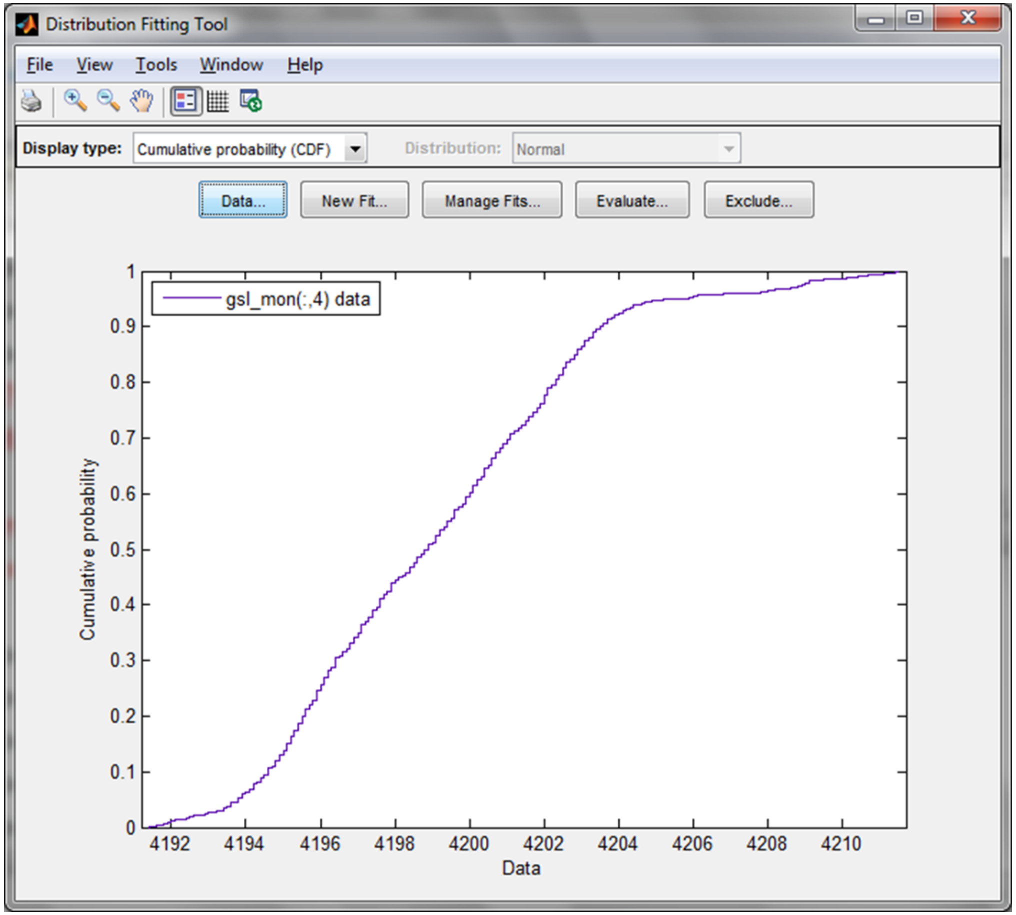
What is the probability of getting blackjack twice: Pr{ ace ∩ 10-K} * Pr{ ace ∩ 10-K} =

Now, play at least 20 hands of blackjack with 3 other people Summarize in a table below your own relative frequencies (a/n) of getting an ace, 10-K, 2-9, blackjack, and two blackjacks. Who in your group was really lucky?

	n	a	a/n
ace			
10-K			
2-9			
blackjack			
Two blackjacks			

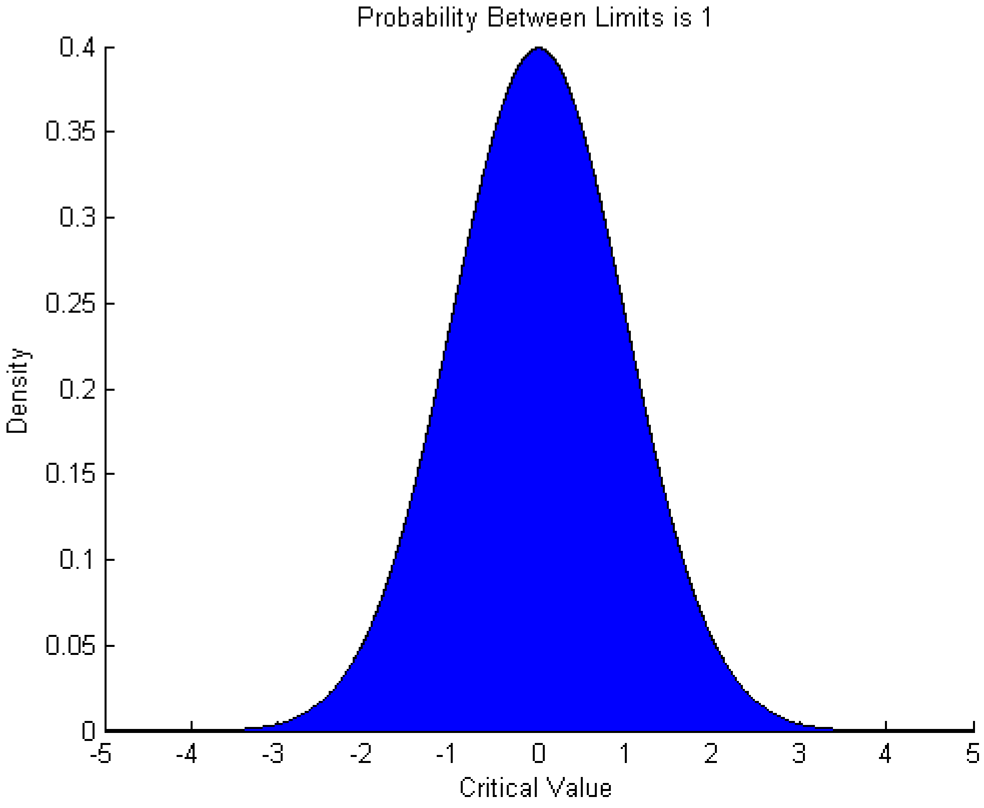
Empirical vs. Parametric Distributions

- Empirical distributions derived from a sample of a population



Empirical vs. Parametric Distributions

- Parameteric distributions:
  - Theoretical approach to define populations with known properties
  - Can be defined by a function with couple parameters and assumption that population composed of random events



Discrete Uniform Distribution

- 6 sided die
- Probability density function:
- f(1,2,3,4,5, or 6) = 1/6
- f(x) = 1/N;
  - x- one of the sides of the die, random variable
  - N- total possible values
  - f(x)dx incremental contribution to total probability

$$\int_1^6 f(x)dx = 1$$

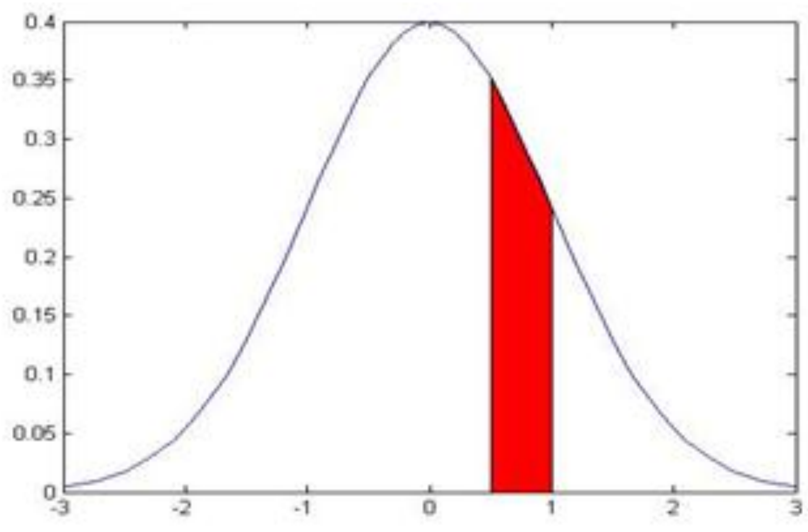
Cumulative Density Function of Discrete Uniform Variable

- F(X)- total probability below a threshhold
- F(X) = x * 1/N
- F(1) = 1/6
- F(2) = 2/6 …
- X(F) – quantile function- value of random variable corresponding to particular cumulative probability
- X(50%) = 3

$$F(X) = \Pr\{x \le X\} = \int_1^X f(x)dx$$

Random Continuous Variable x

- f(x) probability density function (PDF) for a random continuous variable x
- f(x)dx incremental contribution to total probability

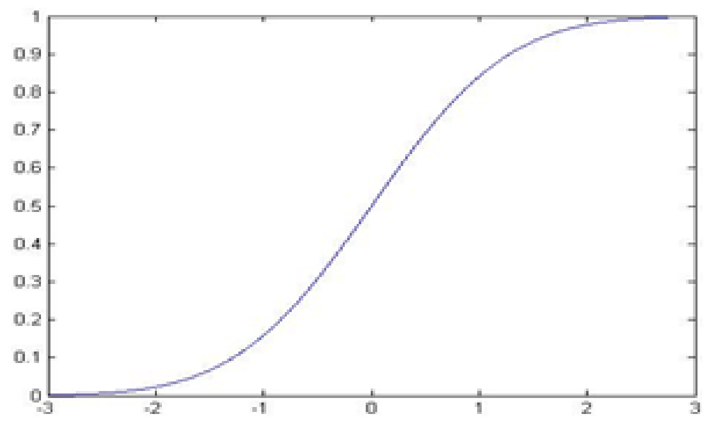


$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Cumulative Density Function of Continuous Variable

- F(X)- total probability below a threshhold
- F(0) = 50%
- F(.66) = 75%
- X(F) – quantile function- value of random variable corresponding to particular cumulative probability
- X(75%) = 0.66

$$F(X) = \Pr\{x \le X\} = \int_{-\infty}^{X} f(x)dx$$



Gaussian Parametric Distribution

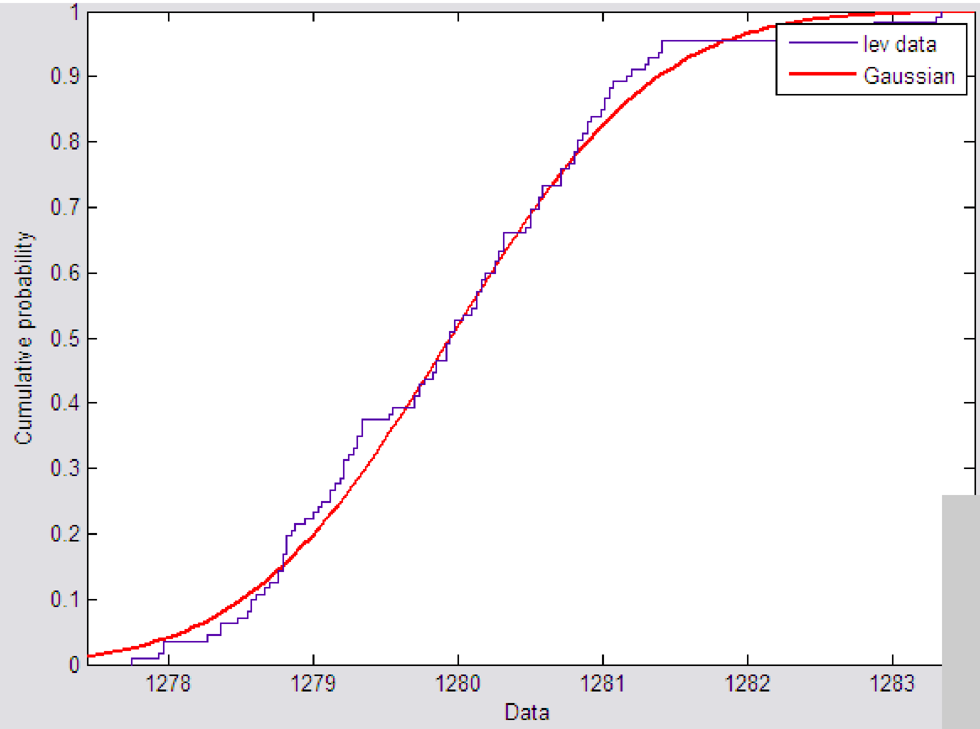
- PDF $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$$
- CDF $$F(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{X} \exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$$

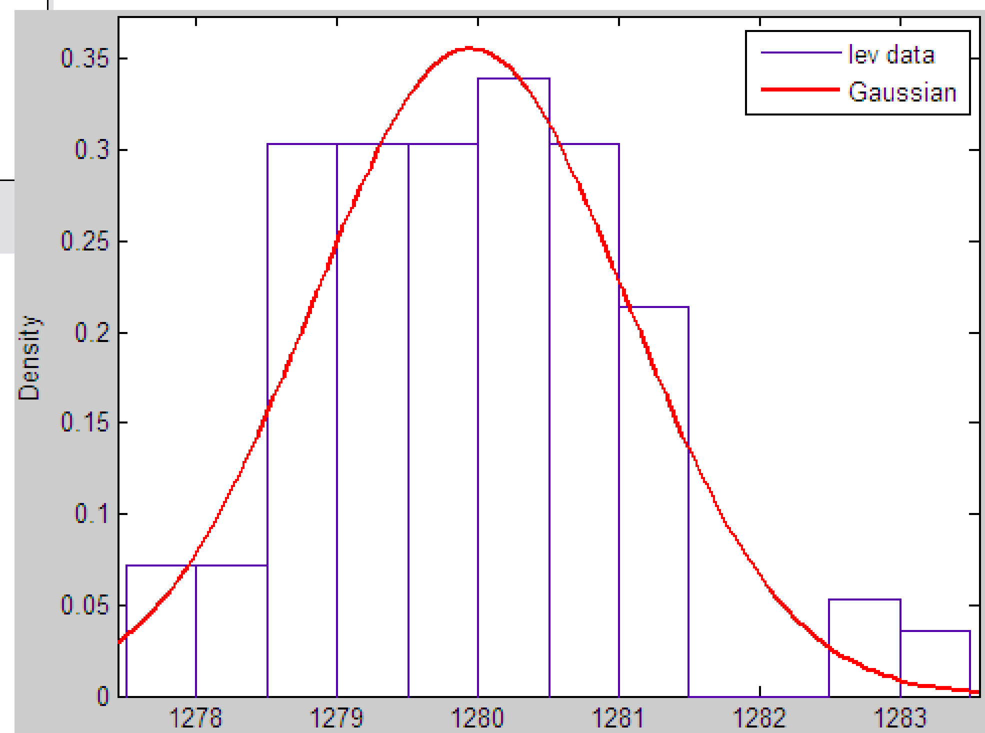
- Two parameters define Gaussian distribution: μ and σ
- Nothing magic or "normal" about the Gaussian distribution- it is a mathematical construct

Using parametric distributions

- Generate an empirical cumulative probability (CDF)
- Use dfittool to see if there is a good match between the empirical CDF and a particular parametric distribution
- Use the parameters from that parametric distribution to estimate the probabilities of values above below a threshhold or extreme events

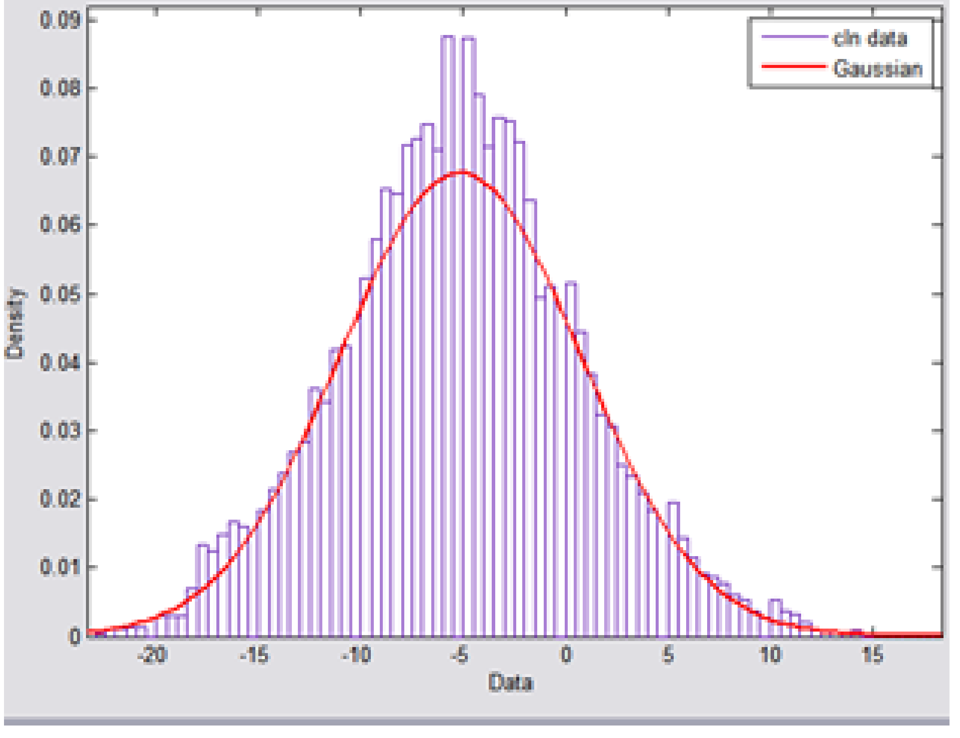
Lake Level

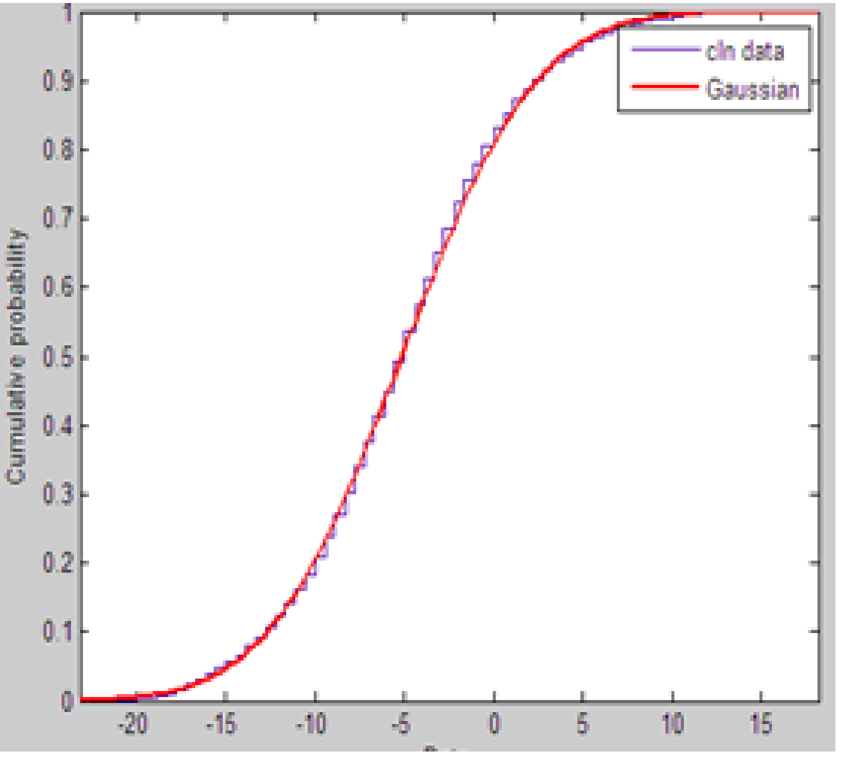




- Fit using sample mean and sample estimate of population std dev

Alta-Collins Temperature

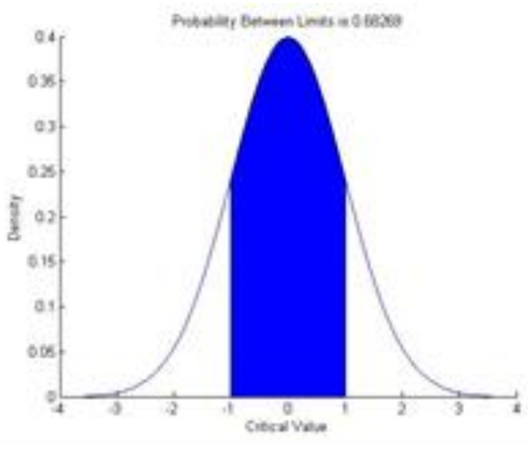


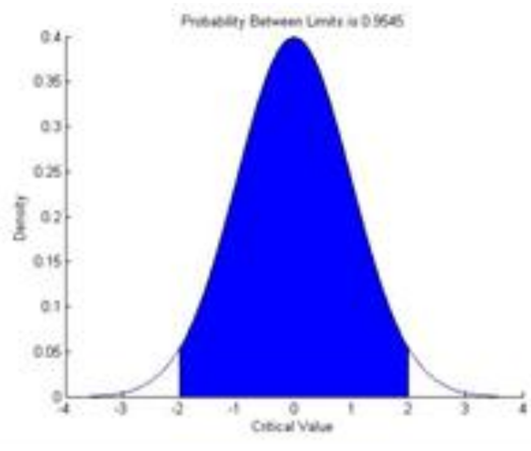


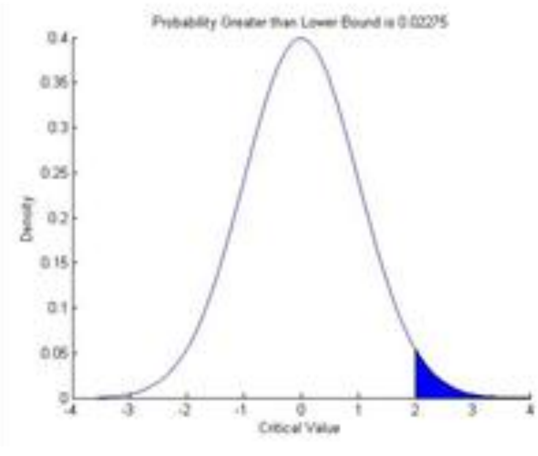
- Mean: -5.1474
- Variance: 34.7724

Using normspec: normal density plot

- 68.3% between -1 and 1
- 95.% between -2 and 2
- 2.3% of time variable explained by Gaussian distribution > 2 std dev of mean

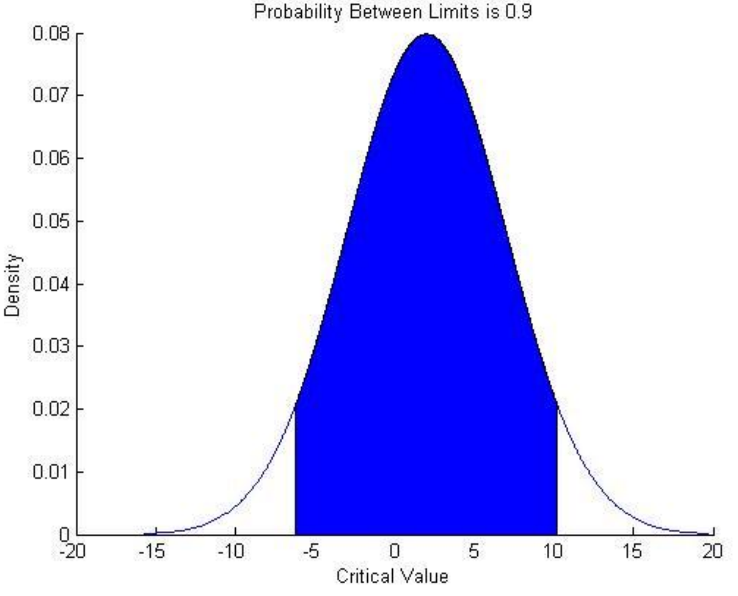






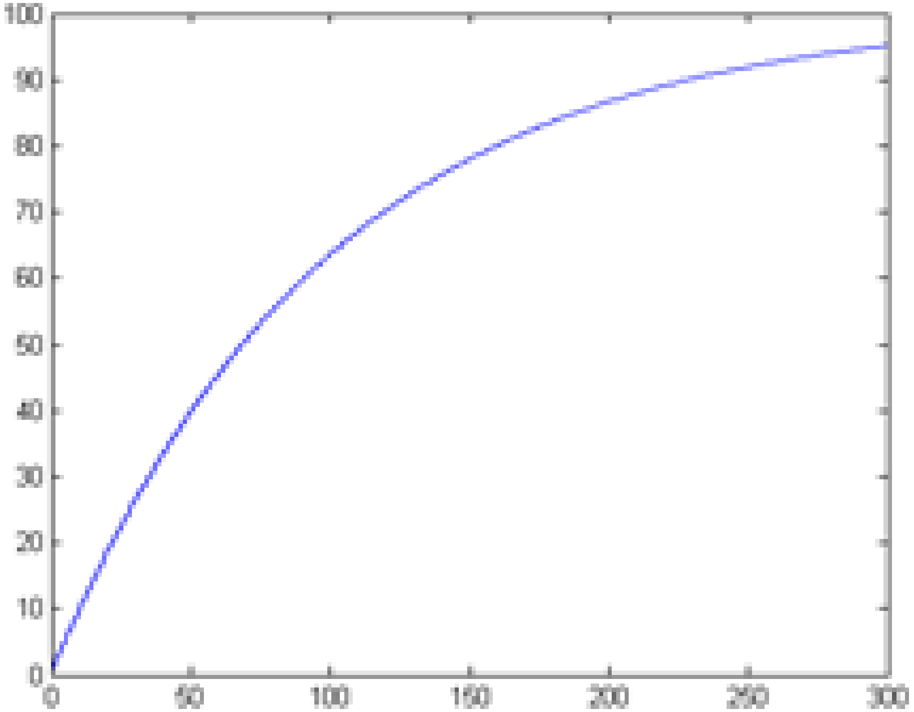
Using norminv: quantile function

- norminv([0.05,0.95],2,5)
- 90% of total variance between -6.2243 10.2243
- normspec([-6.2243,10.2243],2,5)



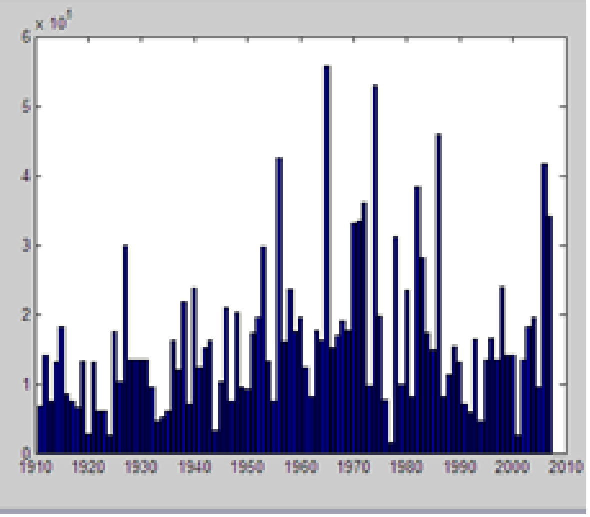
Geometric Distribution

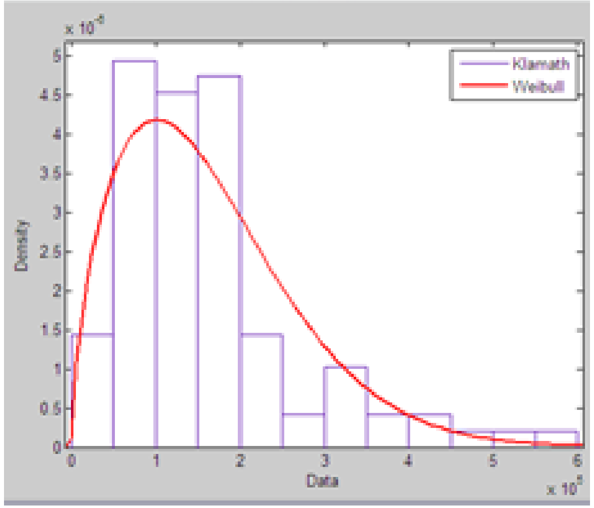
- Estimating how likely rare events can happen by chance
- Pr{0.01}- probability of a 1 in 100 year event
- geocdf(x,0.01)- probability for the next event to happen in 1, 10, 30, 100, 200, 300 years
- 63% chance in next 100 years
- 12% chance not until 200 years

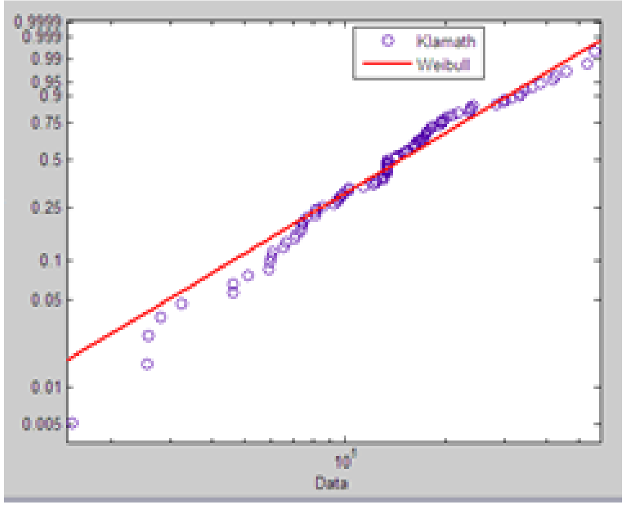


Klamath River Streamflow

- Weibull parametric fit

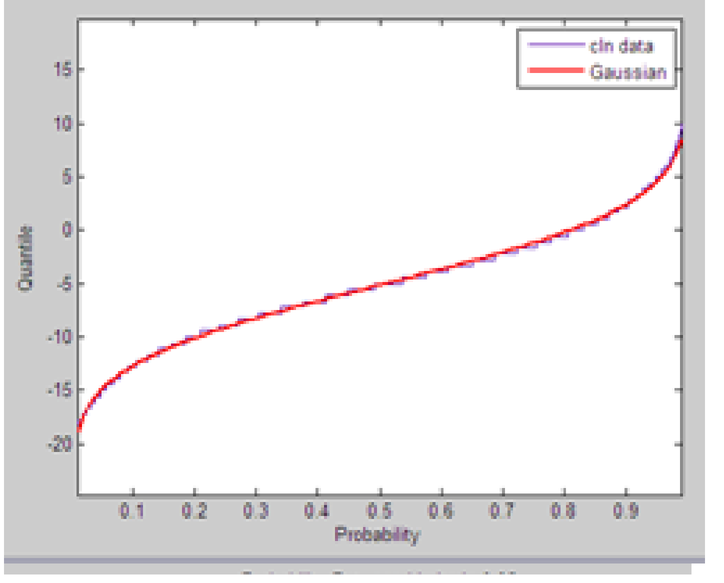






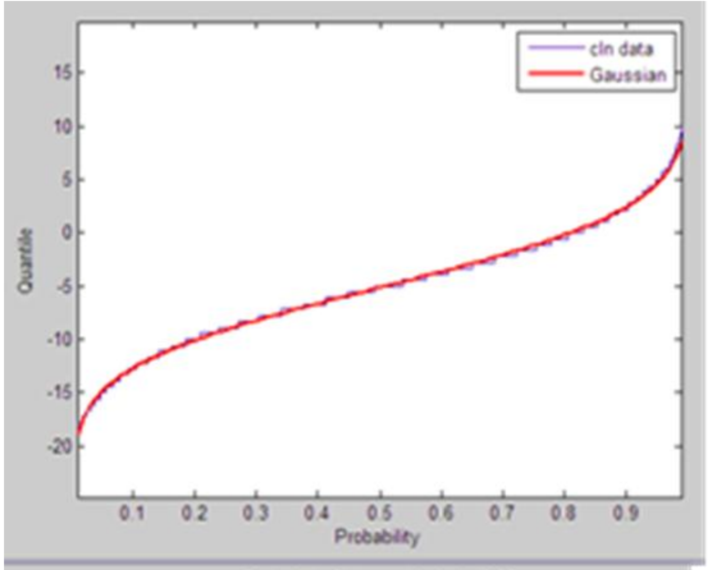
Hypothesis Testing

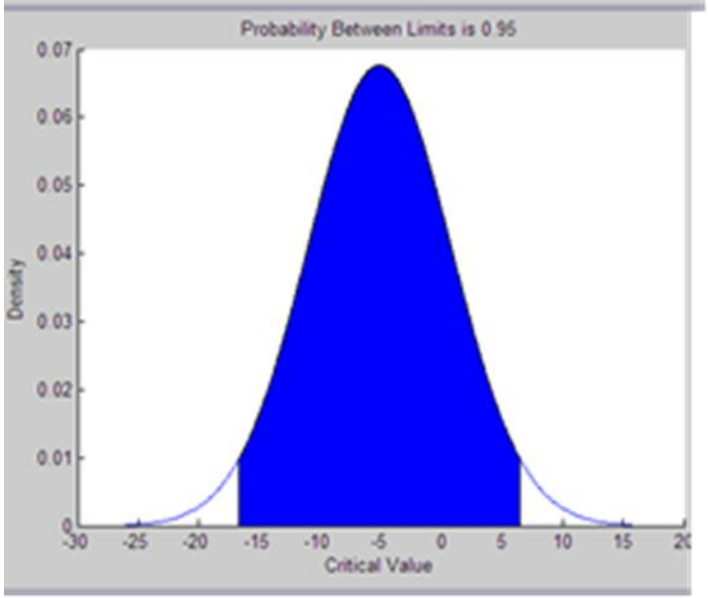
- Alta temperature:
- Empirically: probability of temperature less than -15 is low
- Empirical estimates:
  - Mean= -5.1C
  - Std dev = 5.9C
- What are chances of getting temp of -20 IF this was a population of random numbers with that mean and std dev?



Null hypothesis

- Null hypothesis: Temp of -20C does not differ significantly from mean of -5.1C
- 95% of time, random value would be within -16 and 6C
- So 5% of time, random value would be outside this range
- REJECT the null hypothesis accepting a 5% risk that we are rejecting the null hypothesis incorrectly
- If null hypothesis: Temp of -15C does not differ significantly from mean of -5.1C
- CANNOT reject the null hypothesis since 95% of the time the value could be within -16 and 6C





Collins: Confidence Intervals