Class #10 Wednesday 16 February 2011

What did we cover last time?

- One point correlation map, teleconnection map
- Parametric distributions (why? compactness, smoothing, & extrapolation)
- Parameters versus Statistics
- ~~Discrete distributions~~ Next Time
  - ~~Binomial (N and p)~~
  - ~~Geometric~~
  - ~~Negative Binomial~~
  - ~~Poissons~~

3.3 Statistical Expectations (Ref: Wilkes 4.3)

⁕ **Expected value** of a random variable is the probability-weighted average of that variable.
⁕ E[X], tied to distribution (since provides weights), So for Binomial distribution

$$E[X] = \sum_x x \Pr\{X = x\} = \sum_{x=0}^{N} x \binom{N}{x} p^x (1-p)^{N-x}, x = 0,1,2,3,...N \quad (4.3)$$

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

⁕ Special significance =>E[X] is mean of distribution or denoted μ.

3.3 Statistical Expectations (Ref: Wilkes 4.3)

⁕ Expected value can be evaluated
⁕ Rules in 4.14, constant, multiplicative...
⁕ Evaluated E[X] for our discrete distributions in Table 4.4
⁕ Expectation of $g(x)=(x-\mu)^2$ gives variance Board

$$E[c] = c \quad (4.14a)$$

$$E[c\, g_1(x)] = c\, E[g_1(x)] \quad (4.14b)$$

$$E\left[\sum_{j=1}^{J} g_j(x)\right] = \sum_{j=1}^{J} E[g_j(x)], \quad (4.14c)$$

TABLE 4.4 Expected values (means) and variances for the four discrete probability distribution functions described in Section 4.2, in terms of their distribution parameters.

Distribution	Probability Distribution Function	$\mu = E[X]$	$\sigma^2 = Var[X]$
Binomial	Equation 4.1	Np	Np(1 − p)
Geometric	Equation 4.5	1/p	$(1-p)/p^2$
Negative Binomial	Equation 4.6	k(1 − p)/p	$k(1-p)/p^2$
Poisson	Equation 4.11	μ	μ

3.3 Statistical Expectations (Ref: Wilkes 4.3)

⁕ Example 4.5 Expected Value for Binomial Distribution, N=3 and p=0.5
⁕ X, 0-3 heads example, μ=Np, paid $\$X^2$, average payout would be $E[\$X^2]=\$3.00$

TABLE 4.5 Binomial probabilities for $N = 3$ and $p = 0.5$, and the construction of the expectations $E[X]$ and $E[X^2]$ as probability-weighted averages.

X	Pr(X = x)	x • Pr(X = x)	x^2 • Pr(X = x)
0	0.125	0.000	0.000
1	0.375	0.375	0.375
2	0.375	0.750	1.500
3	0.125	0.375	1.125
		E[X] = 1.500	$E[X^2]$ = 3.000

3.4 Continuous Distributions (Ref: 4.4)

❋ Atmos variables can take on a continuum of values even though measured discretely. We will over a few relevant such distributions.

❋(A) Distribution Functions and Expected Values

❋ Math a bit different than with discrete.

❋ Discontinuous -> Probability Distribution Functions for Continuous ->Probability Density Function (PDF)

❋ Integral over any PDF equals 1

$$\int_x f(x)\,dx = 1 \quad (4.17)$$

3.4 Continuous Distributions cont...

❋(A) Distribution Functions ... cont...

❋ Probability proportional to area (not height) in PDF

❋ Meaningful to determine probability of X between 0.5 and 1.5 (Figure 4.3)

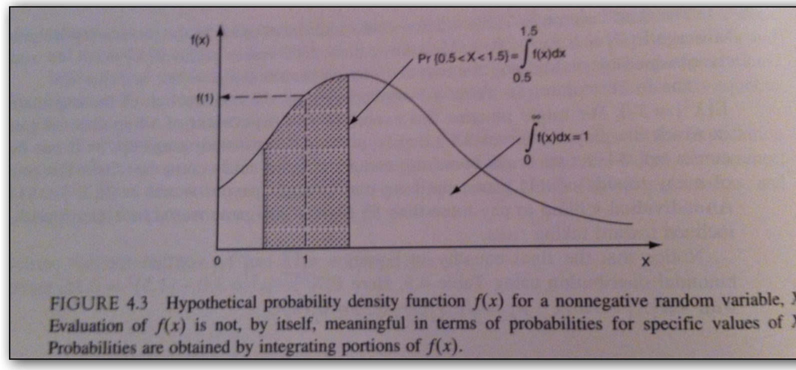


FIGURE 4.3 Hypothetical probability density function $f(x)$ for a nonnegative random variable, X. Evaluation of $f(x)$ is not, by itself, meaningful in terms of probabilities for specific values of X. Probabilities are obtained by integrating portions of $f(x)$.

3.4 Continuous Distributions cont...

❋(A) Distribution Functions ... cont...

❋ Cumulative Distribution Function (CDF) - integral of the PDF up to a certain value

$$F(x) = \Pr\{X \le x\} = \int_{X \le x} f(x)\,dx = 1 \quad (4.18)$$

$0 \le F(x) \le 1$

❋ Inverse Cumulative Distribution Function (CDF) - gives value of random variable corresponding to a particular cumulative probability denoted as $F^{-1}(p)=x(F)$ (4.19), gives data quantile associated with particular probability.

❋ Statistical expectations also defined for continuous distributions (Table 4.6)

3.4 Continuous Distributions cont...

❋(B) Gaussian Distribution

❋ AKA normal distribution, very popular! WHY?

❋ **Central Limit Theorem**, powerful theoretical result, describes the characteristics of the "population of the means" which has been created from the means of an infinite number of random population samples of size (N), all of them drawn from a given "parent population". BOARD

❋ CLT predicts (regardless distribution parent pop):

[1] The mean of the population of means is always equal to the mean of the parent population from which the population samples were drawn.

[2] The standard deviation of the population of means is always equal to the standard deviation of the parent population divided by the square root of the sample size (N).

[3] [And the most amazing part!!] The distribution of means will increasingly approximate a normal distribution as the size N of samples increases.

APPLET Demonstration http://www.chem.uoa.gr/applets/appletcentrallimit/appl_centrallimit2.html

3.4 Continuous Distributions cont...

❋(B) Gaussian Distribution cont...

❋ How large a sample size for CLT to apply? Not clear. Depends on parent population distribution.

❋ Ex: Sum daily temperatures to get monthly mean. Average daily mean Temperatures (avg of max and min).

❋ Ex: Sum daily precipitation values (precip distribution is skewed to the right, many zeros and fewer small values) BOARD, sum more symmetrical than daily.

❋ PDF for a Gaussian distribution, Bell curve shape

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty \quad (4.23)$$

3.4 Continuous Distributions cont...

❋(B) Gaussian Distribution cont...

❋ Need to fit the two parameters sample mean μ and standard deviation σ. How? Method of Moments (calculate moments based on sample)

❋ Once you have the parameters, can integrate Eq 4.23 to get probabilities. Integration needs to be done numerically can't be done analytically or use values in Table B1 (Standardized values)

❋ Use Table B1, need to standardize your parameters. μ =0 & σ =1, standardized PDF is:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] \quad (4.24), \; z = \frac{x - \bar{x}}{s}$$

Do at home: Review Example 4.6 Wilks

3.4 Continuous Distributions cont...

❋(B) Gaussian Distribution cont...

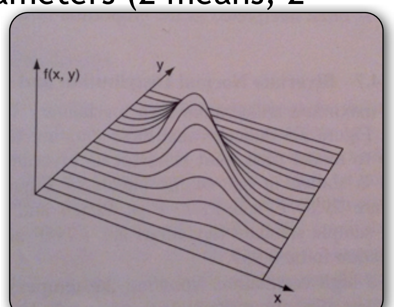
❋ Transform the data in a fancier way to make Gaussian...

❋ Gaussian used so often because it can be generalized to higher dimensions, simplest ex: bivariate Gaussian distribution

❋ EQ 4.33 and 4.34, 5 parameters (2 means, 2 stdevs and a correlation)

❋ Peak at μ's

❋ Probabilities of joint outcomes, integrate PDF



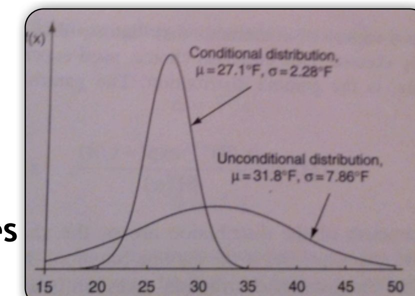
3.4 Continuous Distributions cont...

❋(B) Gaussian Distribution cont...

❋ Conditional distribution (useful property): Note curves in Fig 4.5 are each a gaussian so for the distribution of x given a particular value of y, the Gaussian density function f(x|Y=y) has parameters:

$$\mu_{x|y} = \mu_x + \rho\sigma_x \frac{(y-\mu_y)}{\sigma_y} \quad , \sigma_{x|y} = \sigma_x\sqrt{1-\rho^2} \, (4.37)$$

❋ Conditional $\mu_{x|y}$ larger than μ_x if y larger than mean and rho is positive (vise versa). Conditional stdev smaller. **Knowing something about y reduces uncertainty in x.**



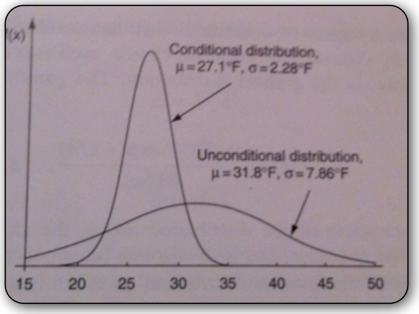
3.4 Continuous Distributions cont...

❋(B) Gaussian Distribution cont...

❋ Conditional distribution (useful property): Note curves in Fig 4.5 are each a gaussian so for the distribution of x given a particular value of y, the Gaussian density function f(x|Y=y) has parameters:

$$\mu_{x|y} = \mu_x + \rho\sigma_x \frac{(y-\mu_y)}{\sigma_y} \quad , \sigma_{x|y} = \sigma_x\sqrt{1-\rho^2} \quad (4.37)$$

❋ Conditional $\mu_{x|y}$ larger than μ_x if y larger than mean and rho is positive (vise versa). Conditional stdev smaller. **Knowing something about y reduces uncertainty in x.** EX:Board



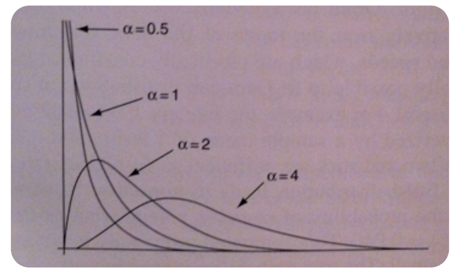
3.4 Continuous Distributions cont...

❋(C) Gamma Distribution

❋Useful for skewed distributions (ex: precip.) so want a function that is bounded on left by zero. The Gamma distribution PDF is:

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp(-a/\beta)}{\beta\Gamma(\alpha)}, \quad x,\alpha,\beta > 0 (4.38)$$

❋ α is shape parameter and β is scale parameter



Gamma distribution density function for different values of alpha. **Bigger alpha ==> less skewness**, Fig 4.7

Methods for α and β

Do at home: Review Example 4.8 Wilks

3.4 Continuous Distributions cont...

❋(D) Gamma Distribution

❋

❋

3.2 Discrete Distributions (Ref: Wilkes 4.2)

- (C) Negative Binomial Distribution
  - Related to geometric distribution, parm p & k
  - Γ function, evaluate numerically

$$PR\{X=x\}=\frac{\Gamma(k+x)}{x!\Gamma(k)}p^k(1-p)^x, x=0,1,2,\ldots(4.6), \Gamma(k+1)=k\Gamma(k)\,(4.8)$$

- X is the number of failures before observing the kth success, so x+k is the waiting time for the kth success.
- Ex 4.3

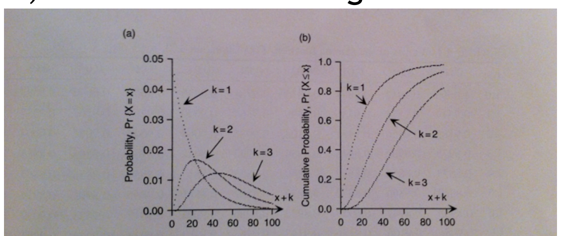


FIGURE 4.1 Probability distribution functions (a), and cumulative probability distribution functions (b), for the waiting time $x+k$ years for Cayuga Lake to freeze k times, using the Negative Binomial distribution, Equation 4.6.

- (D) Poisson Distribution
- Describes a number of discrete events occurring in a sequence. Ex: The occurrence of Atlantic Hurricanes during a particular season.
- Param μ, average occurrence rate, intensity

$$\Pr\{X=x\}=\frac{\mu^x e^{-\mu}}{x!}, x=0,1,2,\ldots(4.11)$$

- Use methods of moments to estimate μ, (mean occurrences per unit time).
- x is number of events, grows small if this number is big, makes sense.

- (D) Poisson Distribution
- Example 4.4, During 30 years there were 138 tornados, so μ is 138/30=4.6 tornados per year.

TABLE 4.3 Numbers of tornados reported annually in New York state, 1959–1988.

1959	3	1969	7	1979	3
1960	4	1970	4	1980	4
1961	5	1971	5	1981	3
1962	1	1972	6	1982	3
1963	3	1973	6	1983	8
1964	1	1974	6	1984	6
1965	5	1975	3	1985	7
1966	1	1976	7	1986	9
1967	2	1977	5	1987	6
1968	2	1978	8	1988	5

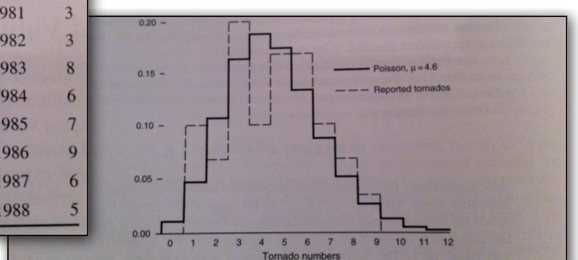


FIGURE 4.2 Histogram of number of tornados reported annually in New York state for 1959–1988 (dashed), and fitted Poisson distribution with $\mu=4.6$ tornados/year (solid).