

UCLA STAT 35
Applied Computational and Interactive Probability

• **Instructor:** Ivo Dinov,
 Asst. Prof. In Statistics and Neurology

• **Teaching Assistant:** Chris Barr

University of California, Los Angeles, Winter 2006
<http://www.stat.ucla.edu/~dinov/>

Stat 35, UCLA, Ivo Dinov Slide 1

Continuous Random Variables and Probability Distributions

Slide 2 Stat 35 UCLA, Ivo Dinov

Continuous Random Variables and Probability Distributions

Slide 3 Stat 35 UCLA, Ivo Dinov

Continuous Random Variables

A random variable X is *continuous* if its set of possible values is an entire interval of numbers (If $A < B$, then any number x between A and B is possible).

Slide 4 Stat 35 UCLA, Ivo Dinov

Probability Distribution

Let X be a continuous rv. Then a *probability distribution or probability density function (pdf)* of X is a function $f(x)$ such that for any two numbers a and b ,

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

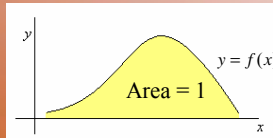
The graph of f is the *density curve*.

Slide 5 Stat 35 UCLA, Ivo Dinov

Probability Density Function

For $f(x)$ to be a pdf

1. $f(x) > 0$ for all values of x .
2. The area of the region between the graph of f and the x – axis is equal to 1.



Slide 6 Stat 35 UCLA, Ivo Dinov

Probability Density Function

$P(a \leq X \leq b)$ is given by the area of the shaded region.

Slide 7 Stat 35 UCLA, Jon Dineen

Continuous RV's

- A RV is **continuous** if it can take on any real value in a non-trivial interval $(a; b)$.
- PDF, probability density function**, for a cont. RV, Y , is a non-negative function $p_Y(y)$, for any real value y , such that for each interval $(a; b)$, the probability that Y takes on a value in $(a; b)$, $P(a < Y < b)$ equals the area under $p_Y(y)$ over the interval $(a; b)$.

Slide 8 Stat 35 UCLA, Jon Dineen

Convergence of density histograms to the PDF

- For a **continuous** RV the density histograms converge to the PDF as the size of the bins goes to zero.

AdditionalInstructorAids\BirthdayDistribution_1978_systat.SYD

Slide 9 Stat 35 UCLA, Jon Dineen

Convergence of density histograms to the PDF

- For a **continuous** RV the density histograms converge to the PDF as the size of the bins goes to zero.

Slide 10 Stat 35 UCLA, Jon Dineen

Uniform Distribution

A continuous rv X is said to have a **uniform distribution** on the interval $[A, B]$ if the pdf of X is

$$f(x; A, B) = \begin{cases} \frac{1}{B - A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

UniformEstimateExperiment

Slide 11 Stat 35 UCLA, Jon Dineen

Probability for a Continuous rv

If X is a continuous rv, then for any number c , $P(x = c) = 0$. For any two numbers a and b with $a < b$,

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X < b) \end{aligned}$$

Slide 12 Stat 35 UCLA, Jon Dineen

Cumulative Distribution Functions and Expected Values

Slide 13 Stat 35 UCLA, Jon Dineen

The Cumulative Distribution Function

The cumulative distribution function, $F(x)$ for a continuous rv X is defined for every number x by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y)dy$$

For each x , $F(x)$ is the area under the density curve to the left of x .

Slide 14 Stat 35 UCLA, Jon Dineen

Using $F(x)$ to Compute Probabilities

Let X be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number a ,

$$P(X > a) = 1 - F(a)$$

and for any numbers a and b with $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a)$$

Slide 15 Stat 35 UCLA, Jon Dineen

Obtaining $f(x)$ from $F(x)$

If X is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every number x for which the derivative $F'(x)$ exists, $F'(x) = f(x)$.

Slide 16 Stat 35 UCLA, Jon Dineen

Percentiles

Let p be a number between 0 and 1. The *(100p)th percentile* of the distribution of a continuous rv X denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy$$

Slide 17 Stat 35 UCLA, Jon Dineen

Median

The *median* of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile. So $\tilde{\mu}$ satisfies $0.5 = F(\tilde{\mu})$. That is, half the area under the density curve is to the left of $\tilde{\mu}$.

Slide 18 Stat 35 UCLA, Jon Dineen

Expected Value

The *expected* or *mean value* of a continuous rv X with pdf $f(x)$ is

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Slide 19 Stat 35 UCLA, Jon Dineen

Expected Value of $h(X)$

If X is a continuous rv with pdf $f(x)$ and $h(x)$ is any function of X , then

$$E[h(x)] = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) \cdot f(x) dx$$

Slide 20 Stat 35 UCLA, Jon Dineen

Variance and Standard Deviation

The *variance* of continuous rv X with pdf $f(x)$ and mean μ is

$$\begin{aligned} \sigma_X^2 = V(x) &= \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx \\ &= E[(X - \mu)^2] \end{aligned}$$

The *standard deviation* is $\sigma_X = \sqrt{V(x)}$.

Slide 21 Stat 35 UCLA, Jon Dineen

Short-cut Formula for Variance

$$V(X) = E(X^2) - [E(X)]^2$$

Slide 22 Stat 35 UCLA, Jon Dineen

The Normal Distribution

A continuous rv X is said to have a normal distribution with parameters μ and σ , where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad -\infty < x < \infty$$

Slide 23 Stat 35 UCLA, Jon Dineen

Slide 24 Stat 35 UCLA, Jon Dineen

Standard Normal Distributions

The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called a *standard normal distribution*. The random variable is denoted by Z . The pdf is

$$f(z; 0, 1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

The cdf is

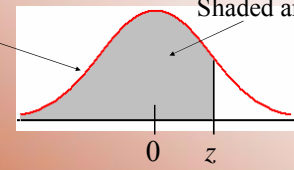
$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z f(y; 0, 1) dy$$

Slide 25

Stat 35 UCLA, Jon Dineen

Standard Normal Cumulative Areas

Standard normal curve



Slide 26

Stat 35 UCLA, Jon Dineen

Standard Normal Distribution

Let Z be the standard normal variable. Find (from table)

a. $P(Z \leq 0.85)$

Area to the left of 0.85 = 0.8023

b. $P(Z > 1.32)$

$$1 - P(Z \leq 1.32) = 0.0934$$

Slide 27

Stat 35 UCLA, Jon Dineen

c. $P(-2.1 \leq Z \leq 1.78)$

Find the area to the left of 1.78 then subtract the area to the left of -2.1.

$$= P(Z \leq 1.78) - P(Z \leq -2.1)$$

$$= 0.9625 - 0.0179$$

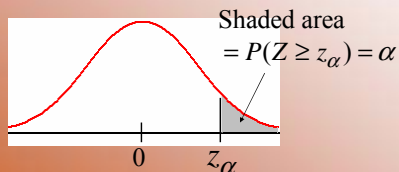
$$= 0.9446$$

Slide 28

Stat 35 UCLA, Jon Dineen

z_α Notation

z_α will denote the value on the measurement axis for which the area under the z curve lies to the right of z_α .



Slide 29

Stat 35 UCLA, Jon Dineen

Ex. Let Z be the standard normal variable. Find z if

a. $P(Z < z) = 0.9278$.

Look at the table and find an entry = 0.9278 then read back to find

$$z = 1.46.$$

b. $P(-z < Z < z) = 0.8132$

$$P(z < Z < -z) = 2P(0 < Z < z)$$

$$= 2[P(z < Z) - \frac{1}{2}]$$

$$= 2P(z < Z) - 1 = 0.8132$$

$$P(z < Z) = 0.9066$$

$$z = 1.32$$

Slide 30

Stat 35 UCLA, Jon Dineen

Nonstandard Normal Distributions

If X has a normal distribution with mean μ and standard deviation σ , then

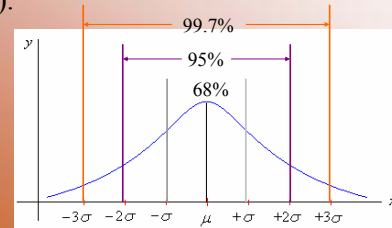
$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution.

Slide 31 Stat 35 UCLA, Jon Dineen

Normal Curve

Approximate percentage of area within given standard deviations (empirical rule).



Slide 32 Stat 35 UCLA, Jon Dineen

Ex. Let X be a normal random variable with $\mu = 80$ and $\sigma = 20$. Find $P(X \leq 65)$.

$$\begin{aligned} P(X \leq 65) &= P\left(Z \leq \frac{65 - 80}{20}\right) \\ &= P(Z \leq -0.75) \\ &= 0.2266 \end{aligned}$$

BivariateNormalExperiment

Slide 33 Stat 35 UCLA, Jon Dineen

Ex. A particular rash shown up at an elementary school. It has been determined that the length of time that the rash will last is normally distributed with $\mu = 6$ days and $\sigma = 1.5$ days.

Find the probability that for a student selected at random, the rash will last for between 3.75 and 9 days.



Slide 34 Stat 35 UCLA, Jon Dineen

$$\begin{aligned} P(3.75 \leq X \leq 9) &= P\left(\frac{3.75 - 6}{1.5} \leq Z \leq \frac{9 - 6}{1.5}\right) \\ &= P(-1.5 \leq Z \leq 2) \\ &= 0.9772 - 0.0668 \\ &= 0.9104 \end{aligned}$$

Slide 35 Stat 35 UCLA, Jon Dineen

Percentiles of an Arbitrary Normal Distribution

$$(100p)\text{th percentile for normal } (\mu, \sigma) = \mu + \left[(100p)\text{th for standard normal} \right] \cdot \sigma$$

Slide 36 Stat 35 UCLA, Jon Dineen

Normal Approximation to the Binomial Distribution

Let X be a binomial rv based on n trials, each with probability of success p . If the binomial probability histogram is not too skewed, X may be approximated by a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$.

$$P(X \leq x) \approx \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

Slide 37

Stat 35 UCLA, Jon Dineen

Ex. At a particular small college the pass rate of Intermediate Algebra is 72%. If 500 students enroll in a semester determine the probability that at least 375 students pass.

$$\mu = np = 500(.72) = 360$$

$$\sigma = \sqrt{npq} = \sqrt{500(.72)(.28)} \approx 10$$

$$P(X \leq 375) \approx \Phi\left(\frac{375.5 - 360}{10}\right) = \Phi(1.55) = 0.9394$$

Slide 38

Stat 35 UCLA, Jon Dineen

Normal approximation to Binomial

- Suppose $Y \sim \text{Binomial}(n, p)$
- Then $Y = Y_1 + Y_2 + Y_3 + \dots + Y_n$, where
 - $Y_k \sim \text{Bernoulli}(p)$, $E(Y_k) = p$ & $\text{Var}(Y_k) = p(1-p) \rightarrow$
 - $E(Y) = np$ & $\text{Var}(Y) = np(1-p)$, $\text{SD}(Y) = (np(1-p))^{1/2}$
 - **Standardize Y:**
 - $Z = (Y - np) / (np(1-p))^{1/2}$
 - By CLT $\rightarrow Z \sim N(0, 1)$. So, $Y \sim N(np, (np(1-p))^{1/2})$
- **Normal Approx to Binomial is reasonable when $np \geq 10$ & $n(1-p) > 10$** (p & $(1-p)$ are NOT too small relative to n).

Slide 39

Stat 35 UCLA, Jon Dineen

Normal approximation to Binomial – Example

- **Roulette wheel investigation:**
- Compute $P(Y \geq 58)$, where $Y \sim \text{Binomial}(100, 0.47)$ –
 - The proportion of the Binomial(100, 0.47) population having more than 58 reds (successes) out of 100 roulette spins (trials).
 - Since $np = 47 \geq 10$ & $n(1-p) = 53 > 10$ Normal approx is justified.
- $Z = (Y - np) / \text{Sqrt}(np(1-p)) = \frac{58 - 100 \cdot 0.47}{\text{Sqrt}(100 \cdot 0.47 \cdot 0.53)} = 2.2$
- $P(Y \geq 58) \leftrightarrow P(Z \geq 2.2) = 0.0139$
- True $P(Y \geq 58) = 0.177$, using SOCR (demo!)
- Binomial approx useful when no access to SOCR avail.

Roulette has 38 slots
18 red 18 black 2 neutral

Slide 40

Stat 35 UCLA, Jon Dineen

Normal approximation to Poisson

- Let $X_1 \sim \text{Poisson}(l)$ & $X_2 \sim \text{Poisson}(m) \rightarrow X_1 + X_2 \sim \text{Poisson}(l+m)$
- Let $X_1, X_2, X_3, \dots, X_k \sim \text{Poisson}(l)$, and independent,
- $Y_k = X_1 + X_2 + \dots + X_k \sim \text{Poisson}(kl)$, $E(Y_k) = \text{Var}(Y_k) = kl$.
- The random variables in the sum on the right are **independent** and each has the Poisson distribution with parameter l .
- By CLT the distribution of the standardized variable $(Y_k - kl) / (kl)^{1/2} \rightarrow N(0, 1)$, as k increases to infinity.
- So, for $kl \geq 100$, $Z_k = \{(Y_k - kl) / (kl)^{1/2}\} \sim N(0, 1)$.
- $\rightarrow Y_k \sim N(kl, (kl)^{1/2})$.

Slide 41

Stat 35 UCLA, Jon Dineen

Normal approximation to Poisson – example

- Let $X_1 \sim \text{Poisson}(l)$ & $X_2 \sim \text{Poisson}(m) \rightarrow X_1 + X_2 \sim \text{Poisson}(l+m)$
- Let $X_1, X_2, X_3, \dots, X_{200} \sim \text{Poisson}(2)$, and independent.
- $Y_k = X_1 + X_2 + \dots + X_k \sim \text{Poisson}(400)$, $E(Y_k) = \text{Var}(Y_k) = 400$.
- By CLT the distribution of the standardized variable $(Y_k - 400) / (400)^{1/2} \rightarrow N(0, 1)$, as k increases to infinity.
- $Z_k = (Y_k - 400) / 20 \sim N(0, 1) \rightarrow Y_k \sim N(400, 400)$.
- $P(2 < Y_k < 400) = (\text{std}'z \ 2 \ \& \ 400) =$
- $P((2-400)/20 < Z_k < (400-400)/20) = P(-20 < Z_k < 0) = 0.5$



Slide 42

Stat 35 UCLA, Jon Dineen

Poisson or Normal approximation to Binomial?

- Poisson Approximation (Binomial(n, p_n) → Poisson(λ)):

$$\binom{n}{y} p_n^y (1 - p_n)^{n-y} \xrightarrow[n \times p_n \rightarrow \lambda]{\substack{\text{WHY?} \\ n \rightarrow \infty}} \frac{\lambda^y e^{-\lambda}}{y!}$$
 - n ≥ 100 & p ≤ 0.01 & λ = n p ≤ 20
- Normal Approximation (Binomial(n, p) → N (np, (np(1-p))^{1/2})
 - np ≥ 10 & n(1-p) > 10

Slide 43 Stat 35 UCLA, Ivo Dinov

The Gamma Distribution and Its Relatives

Slide 44 Stat 35 UCLA, Ivo Dinov

The Gamma Function

For α > 0, the *gamma function* Γ(α) is defined by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

Gamma Distribution

Slide 45 Stat 35 UCLA, Ivo Dinov

Gamma Distribution

A continuous rv X has a *gamma distribution* if the pdf is

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where the parameters satisfy α > 0, β > 0. The standard gamma distribution has β = 1.

Slide 46 Stat 35 UCLA, Ivo Dinov

Mean and Variance

The mean and variance of a random variable X having the gamma distribution f(x; α, β) are

$$E(X) = \mu = \alpha\beta \quad V(X) = \sigma^2 = \alpha\beta^2$$

Slide 47 Stat 35 UCLA, Ivo Dinov

Probabilities from the Gamma Distribution

Let X have a gamma distribution with parameters α and β.

Then for any x > 0, the cdf of X is given by

$$P(X \leq x) = F(x; \alpha, \beta) = F\left(\frac{x}{\beta}; \alpha\right)$$

where

$$F(x; \alpha) = \int_0^x \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)} dy$$

Slide 48 Stat 35 UCLA, Ivo Dinov

Exponential Distribution

A continuous rv X has an *exponential distribution* with parameter λ if the pdf is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

ExponentialDistribution

Slide 49 Stat 35 UCLA, Jon Dineen

Mean and Variance

The mean and variance of a random variable X having the exponential distribution

$$\mu = \alpha\beta = \frac{1}{\lambda} \quad \sigma^2 = \alpha\beta^2 = \frac{1}{\lambda^2}$$

Slide 50 Stat 35 UCLA, Jon Dineen

Probabilities from the Gamma Distribution

Let X have a exponential distribution
Then the cdf of X is given by

$$F(x; \lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

Slide 51 Stat 35 UCLA, Jon Dineen

Applications of the Exponential Distribution

Suppose that the number of events occurring in any time interval of length t has a Poisson distribution with parameter αt and that the numbers of occurrences in nonoverlapping intervals are independent of one another. Then the distribution of elapsed time between the occurrences of two successive events is exponential with parameter $\lambda = \alpha$.

Slide 52 Stat 35 UCLA, Jon Dineen

The Chi-Squared Distribution

Let ν be a positive integer. Then a random variable X is said to have a *chi-squared distribution* with parameter ν if the pdf of X is the gamma density with $\alpha = \nu/2$ and $\beta = 2$. The pdf is

$$f(x; \nu) = \begin{cases} \frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

ChiSquareDistribution

Slide 53 Stat 35 UCLA, Jon Dineen

The Chi-Squared Distribution χ^2

The parameter ν is called the *number of degrees of freedom* (df) of X .

Applications:

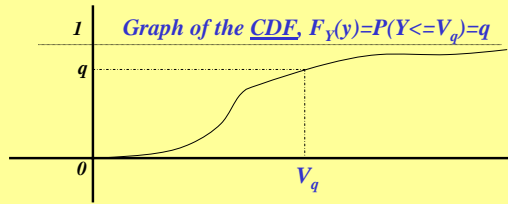
1. F-test = $\chi^2_{(df=\nu_1)} / \chi^2_{(df=\nu_2)}$
2. Sample-Variiances $\sim \chi^2_{(df=\nu)}$

Model Fitting Demo: (SOCR Additional Resources)
<http://socr.stat.ucla.edu/Applets.dir/SOCRCurveFitter.html>

Slide 54 Stat 35 UCLA, Jon Dineen

Identifying Common Distributions – QQ plots

- **Quantile-Quantile** plots indicate how well the model distribution agrees with the data.
- q^{th} quantile, for $0 < q < 1$, is the (data-space) value, V_q , at or below which lies a proportion q of the data.



Slide 55 Stat 35 UCLA, Jon Dineen

Constructing QQ plots

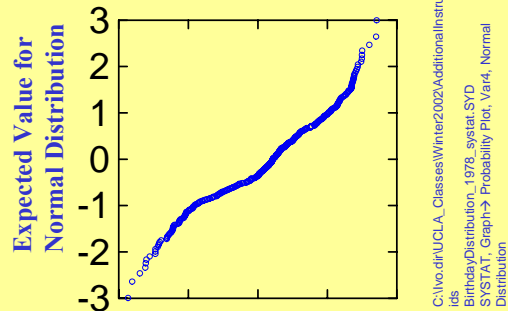
- Start off with data $\{y_1, y_2, y_3, \dots, y_n\}$
- Order statistics $y_{(1)} \leq y_{(2)} \leq y_{(3)} \leq \dots \leq y_{(n)}$
- Compute **quantile rank**, $q_{(k)}$, for each observation, $y_{(k)}$,

$$P(Y \leq q_{(k)}) = (k - 0.375) / (n + 0.250),$$
 where Y is a RV from the (target) model distribution.
- Finally, plot the points $(y_{(k)}, q_{(k)})$ in 2D plane, $1 \leq k \leq n$.
- **Note:** Different statistical packages use slightly different formulas for the computation of $q_{(k)}$. However, the results are quite similar. This is the formulas employed in SAS.
- **Basic idea:** Probability that: $P((\text{model})Y \leq (\text{data})y_{(1)}) \sim 1/n$;
 $P(Y \leq y_{(2)}) \sim 2/n$; $P(Y \leq y_{(3)}) \sim 3/n$; ...

Slide 56 Stat 35 UCLA, Jon Dineen

Example - Constructing QQ plots

- Start off with data $\{y_1, y_2, y_3, \dots, y_n\}$.
- Plot the points $(y_{(k)}, q_{(k)})$ in 2D plane, $1 \leq k \leq n$



Slide 57 Stat 35 UCLA, Jon Dineen

Other Continuous Distributions

Slide 58 Stat 35 UCLA, Jon Dineen

The Weibull Distribution

A continuous rv X has a **Weibull distribution** if the pdf is

$$f(x; \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where the parameters satisfy $\alpha > 0, \beta > 0$.

WeibullDistribution

Slide 59 Stat 35 UCLA, Jon Dineen

Mean and Variance

The mean and variance of a random variable X having the Weibull distribution are

$$\mu = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \quad \sigma^2 = \beta^2 \left\{ \Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right) \right]^2 \right\}$$

Slide 60 Stat 35 UCLA, Jon Dineen

Weibull Distribution

The cdf of a Weibull rv having parameters α and β is

$$F(x; \alpha, \beta) = \begin{cases} 1 - e^{-(x/\beta)^\alpha} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Slide 61 Stat 35 UCLA, Jon Dineen

Lognormal Distribution

A nonnegative rv X has a *lognormal distribution* if the rv $Y = \ln(X)$ has a normal distribution the resulting pdf has parameters μ and σ and is

$$f(x; \mu, \sigma) = \begin{cases} \frac{1}{\sqrt{2\pi}\alpha x} e^{-[\ln(x)-\mu]^2/(2\sigma^2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

RandomVariableExperiment
LogNormalDistribution

Slide 62 Stat 35 UCLA, Jon Dineen

Mean and Variance

The mean and variance of a variable X having the lognormal distribution are

$$E(X) = e^{\mu + \sigma^2/2} \quad V(X) = e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1)$$

Slide 63 Stat 35 UCLA, Jon Dineen

Lognormal Distribution

The cdf of the lognormal distribution is given by

$$\begin{aligned} F(x; \mu, \sigma) &= P(X \leq x) = P[\ln(X) \leq \ln(x)] \\ &= P\left(Z \leq \frac{\ln(x) - \mu}{\sigma}\right) = \Phi\left(\frac{\ln(x) - \mu}{\sigma}\right) \end{aligned}$$

Slide 64 Stat 35 UCLA, Jon Dineen

Beta Distribution

A rv X is said to have a *beta distribution* with parameters $A, B, \alpha > 0$, and $\beta > 0$ if the pdf of X is

$$f(x; \alpha, \beta, A, B) = \begin{cases} \frac{1}{B-A} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \left(\frac{x-A}{B-A}\right)^{\alpha-1} \left(\frac{B-x}{B-A}\right)^{\beta-1} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

BetaDistribution
BetaCoinExperiment BetaEstimateExperiment

Slide 65 Stat 35 UCLA, Jon Dineen

Mean and Variance

The mean and variance of a variable X having the beta distribution are

$$\begin{aligned} \mu &= A + (B - A) \cdot \frac{\alpha}{\alpha + \beta} \\ \sigma^2 &= \frac{(B - A)^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} \end{aligned}$$

Slide 66 Stat 35 UCLA, Jon Dineen

Probability Plots

Slide 67 Stat 35 UCLA, Jon Dinger

Sample Percentile

Order the n -sample observations from smallest to largest. The i th smallest observation in the list is taken to be the $[100(i - 0.5)/n]$ th *sample percentile*.

Slide 68 Stat 35 UCLA, Jon Dinger

Probability Plot

$\left(\begin{array}{l} [100(i - .5)/n] \text{th percentile} \\ \text{of the distribution} \end{array} \right)$ i th smallest sample observation

If the sample percentiles are close to the corresponding population distribution percentiles, the first number will roughly equal the second.

ProbabilityPlotExperiment

Slide 69 Stat 35 UCLA, Jon Dinger

Normal Probability Plot

A plot of the pairs

$\left([100(i - .5)/n] \text{th } z \text{ percentile}, i \text{th smallest observation} \right)$

On a two-dimensional coordinate system is called a normal probability plot. If the drawn from a normal distribution the points should fall close to a line with slope σ and intercept μ .

SOCRCharts WebStat QQ-Normal Plot

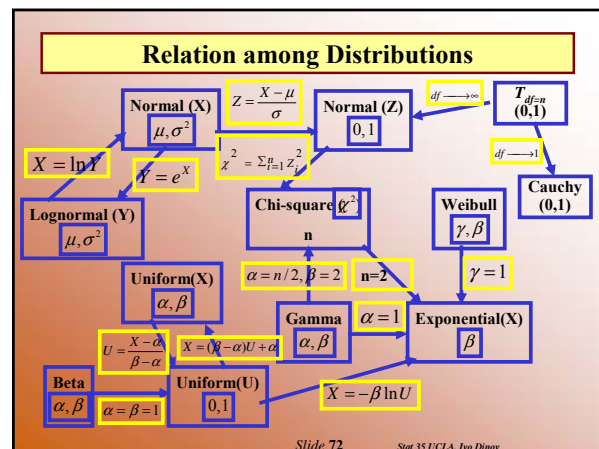
Slide 70 Stat 35 UCLA, Jon Dinger

Beyond Normality

Consider a family of probability distributions involving two parameters θ_1 and θ_2 . Let $F(x; \theta_1, \theta_2)$ denote the corresponding cdf's. The parameters θ_1 and θ_2 are said to *location* and *scale* parameters if

$$F(x; \theta_1, \theta_2) \text{ is a function of } \frac{x - \theta_1}{\theta_2}.$$

Slide 71 Stat 35 UCLA, Jon Dinger



Slide 72 Stat 35 UCLA, Jon Dinger