UCLA STAT 110 A
Applied Probability & Statistics for Engineers

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

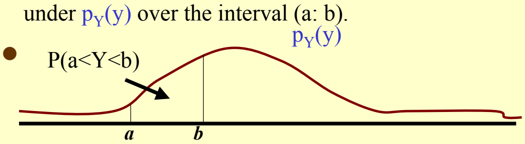
•**Teaching Assistant:** Maria Chang, UCLA Statistics

University of California, Los Angeles, Spring 2003
http://www.stat.ucla.edu/~dinov/

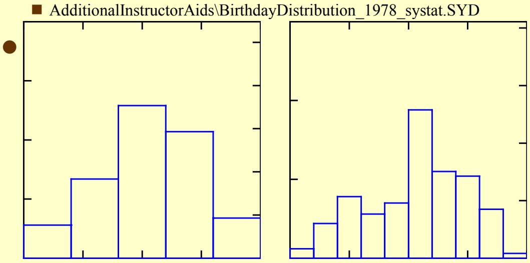
Chapters 4: Continuous Variables, Continuous Probability Density Functions

- Continuous RV's PDF's
- Normal, Gamma, Exponential, χ^2, F, T distributions
- Central Limit Theorem (CLT)

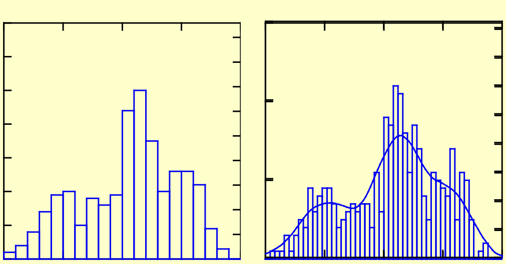
Continuous RV's

- 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).
- 

Convergence of density histograms to the PDF

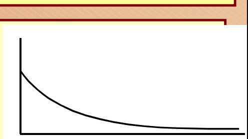
- For a continuous RV the density histograms converge to the PDF as the size of the bins goes to zero.
  - AdditionalInstructorAids\BirthdayDistribution_1978_systat.SYD
- 

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.
- 

Computing Probabilities using PDFs

- $$P(Y \in A) = \int_A p_Y(y)dy$$
  $$p_Y(y) = e^{-y}, y \ge 0$$



(i) Exponential shape

- Example:
  $$P(0 \le Y \le 3) = \int_0^3 p_Y(y)dy =$$
  $$\int_0^3 e^{-y}dy = -\left. e^{-y} \right|_0^3 = 1 - e^{-3} \cong 1$$

CDF (cumulative distribution function)

- $F_Y(y) = P(Y \le y) = \int_{-\infty}^{y} p_Y(y)dy$
- Example:

$$p_Y(y) = e^{-y}, y \ge 0$$

$$F_Y(3) = P(Y \le 3) = \int_0^3 p_Y(y)dy =$$

$$\int_0^3 e^{-y}dy = -\left. e^{-y} \right|_0^3 = 1 - e^{-3} \cong 1$$

Measures of central tendency/variability for Continuous RVs

- Mean $\mu_Y = \int_{-\infty}^{\infty} y \times p_Y(y)dy$
- Variance $\sigma_Y^2 = \int_{-\infty}^{\infty} (y-\mu_Y)^2 \times p_Y(y)dy$
- SD $\sigma_Y = \sqrt{\int_{-\infty}^{\infty} (y-\mu_Y)^2 \times p_Y(y)dy}$

Facts about PDF's of continuous RVs

- Non-negative $p_Y(y) \ge 0, \forall y$
- Completeness $\int_{-\infty}^{\infty} p_Y(y)dy = 1$
- Probability

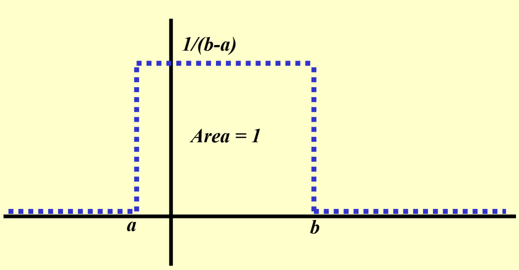
$$P(a < Y < b) = \int_a^b y \times p_Y(y)dy$$

Continuous Distributions

- Uniform distribution
- Normal distribution
- Student's T distribution
- F-distribution
- Chi-squared (χ^2)
- Cauchy's distribution
- Exponential distribution
- Poisson distribution, …

Uniform Distribution

- Uniform Distribution PDF: **Y~Uniform(a;b)** ←→ $p_Y(y)=1/(b-a)$, for each a<=y<=b, and $p_Y(y)=0$, otherwise.



(Continuous) Uniform Distribution

- $X \sim$ Uniform Distribution with parameters α and β if

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & , \quad \alpha < x < \beta \\ 0 & , \quad \text{otherwise} \end{cases}$$

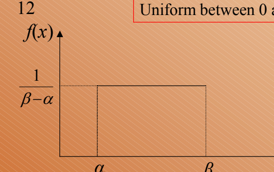
$E(X) = \frac{\alpha+\beta}{2}$, $Var(X) = \frac{(\beta-\alpha)^2}{12}$

• random numbers follow Uniform between 0 and 1

Ex) Uniform, $\alpha = 2, \beta = 7$

(a) $P(X \ge 4) =$

(b) $P(3 < X < 5.5) =$



Uniform Distribution – CDF, mean, variance

- Uniform Distribution CDF:

$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_{a}^{\min(y,b)} \frac{1}{b-a}dx = \left.\frac{x}{b-a}\right|_a^y = \begin{cases} 0, & y < a \\ \frac{y-a}{b-a}, & a \le y \le b \\ 1, & b < y \end{cases}$$

Uniform Distribution – CDF, mean, variance

- Mean:

$$\mu_Y = \int_{-\infty}^{\infty} y p_Y(y)dy = \int_a^b \frac{y}{b-a}dy = \left.\frac{y^2}{2(b-a)}\right|_a^b = \frac{a+b}{2}$$

- Variance:

$$\sigma_Y^2 = \int_{-\infty}^{\infty} (y-\mu_Y)^2 p_Y(y)dy = \int_a^b \frac{(2y-(a+b))^2}{4(b-a)}dy = \frac{(b-a)^2}{12}$$

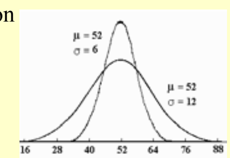
- SD:

$$\sigma_Y = \sqrt{\int_{-\infty}^{\infty} (y-\mu_Y)^2 p_Y(y)dy} = \frac{(b-a)}{\sqrt{12}}$$

Continuous Distributions - Normal

- (General) Normal distribution

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$



- (Standard) Normal distribution (μ=0, σ=1)

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \qquad Z = \frac{Y-\mu}{\sigma}$$

(General) Normal Distribution

- Normal Distribution PDF: **Y~Normal(μ, σ²) ←→**

$$p_Y(y) = \frac{e^{-\frac{(y-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}, \forall -\infty < y < \infty$$

$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_{-\infty}^{y} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}dx$$

Standard Normal (Gaussian) Distribution

- Normal Distribution PDF: **Y~Normal(μ=0, σ²=1) ←→**

$$p_Y(y) = \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}, \forall -\infty < y < \infty$$

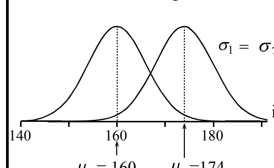
$$F_Y(y) = \int_{-\infty}^{y} p_Y(x)dx = \int_{-\infty}^{y} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx$$

Effects of μ and σ (on the graphs of Normal Distribution)

(a) Changing μ

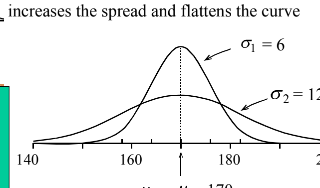
shifts the curve along the axis

Mean is a measure of …
central tendency



(b) Increasing σ

increases the spread and flattens the curve

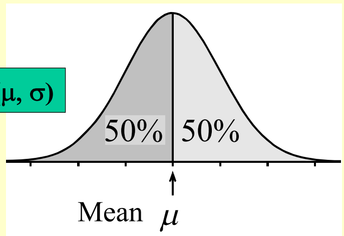


Standard deviation is a measure of …
variability/spread

The Normal distribution density curve

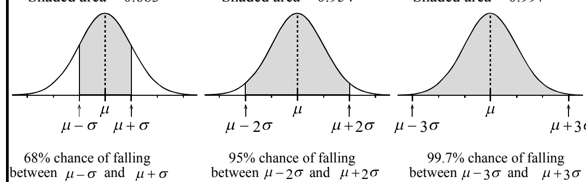
- Is symmetric about the mean! Bell-shaped and unimodal.
- Mean = Median!

N(μ, σ)



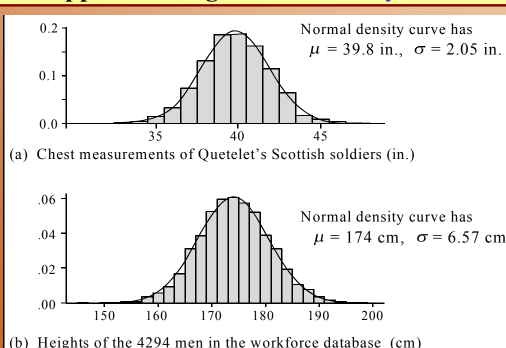
Understanding the standard deviation: σ

Probabilities/areas and numbers of standard deviations for the Normal distribution



NormalCurveInteractive.html

Two standardized histograms with approximating Normal density curve



Basic method for obtaining probabilities

- Sketch a Normal curve, marking the mean and other values of interest.
- Shade the area under the curve that gives the desired probability.
- Devise a way of getting the desired area from lower-tail areas.
- Obtain component lower-tail probabilities from a computer program

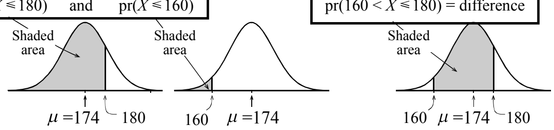
(a) Computing pr(160 < X ≤ 180)

Programs supply

pr(X ≤ 180) and pr(X ≤ 160)

We want

pr(160 < X ≤ 180) = difference



pr(160 < X ≤ 180) = pr(X ≤ 180) − pr(X ≤ 160)

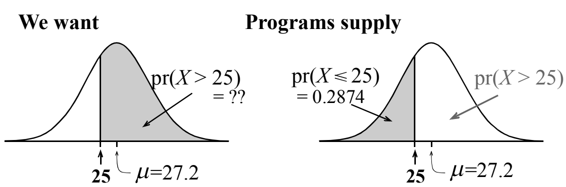
Tabular representation of probabilities

(c) More Normal probabilities (*values obtained from Minitab*)

b	pr(X ≤ b)	a	pr(X ≤ a)	pr(a < X ≤ b) = difference
167.6	0.165	152.4	0.001	0.164
177.8	0.718	167.6	0.165	0.553
177.8	0.718	152.4	0.001	0.717
182.9	0.912	167.6	0.165	0.747

Note: 152.4cm = 5ft, 167.6cm = 5ft 6in., 177.8cm = 5ft 10in., 182.9cm = 6ft

Obtaining an upper-tail probability



Since total area under curve = 1, pr(X > 25) = 1 - pr($X \le$ 25)

Generally, **pr($X > x$) = 1 − pr($X \le x$)**

Continuous Distributions – Student's T

- Student's T distribution [approx. of Normal(0,1)]
  - $Y_1, Y_2, \ldots, Y_N$ IID from a Normal($\mu;\sigma$)
  - Variance σ^2 is unknown
- In 1908, William Gosset (pseudonym Student) derived the exact sampling distribution of the following statistics

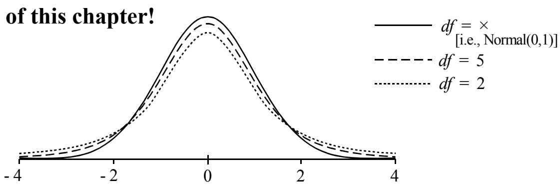
$$T = \frac{\bar{Y} - \mu_Y}{\hat{\sigma}_Y}$$

- T~Student(**df=N-1**), where

$$\hat{\sigma}_Y = \sqrt{\frac{\sum_{k=1}^{N}(Y_k - \bar{Y})^2}{N-1}}$$

Density curves for Student's *t*

We will come back to the T-distribution at the end of this chapter!



Student(df) density curves for various *df*.

Bias and Precision

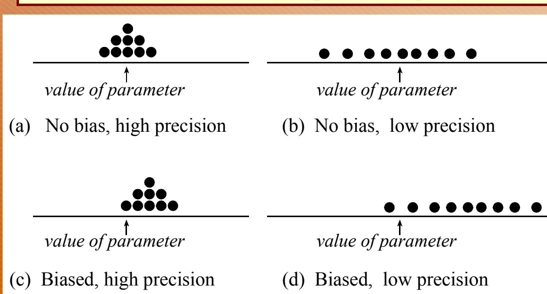
- The bias in an estimator is the distance between between the *center* of the sampling distribution of the estimator and the true value of the parameter being estimated. In math terms, bias = $E(\hat{\Theta}) - \theta$, where theta $\hat{\Theta}$ is the estimator, as a RV, of the true (unknown) parameter θ .
- Example, Why is the sample mean an **unbiased** estimate for the population mean? How about ¾ of the sample mean?

$$E(\hat{\Theta}) - \mu = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) - \mu = 0$$

$$E(\hat{\Theta}) - \mu = E\left(\frac{3}{4}\frac{1}{n}\sum_{k=1}^{n} X_k\right) - \mu = \frac{3}{4}\mu - \mu = \frac{\mu}{4} \ne 0, \quad \text{in general.}$$

Bias and Precision

- The precision of an estimator is a measure of how variable is the estimator in repeated sampling.



Bias and Precision – 2 unbiased estimators of μ

- Suppose: $\{Y_1, Y_2, \ldots\ldots\ldots\ldots, Y_{4N}\}$ IID (μ, σ).
- Let $\bar{Y}_1 = {}^{1}\!/_{3N}\sum_{k=1}^{3N} Y_k; \quad \bar{Y}_2 = {}^{1}\!/_{N}\sum_{k=3N+1}^{4N} Y_k$
- And $\bar{Y} = \frac{\bar{Y}_1 + \bar{Y}_2}{2} \quad \bar{\bar{Y}} = \sum_{k=1}^{4N} Y_k;$
- Which estimator is better? $\bar{Y}$ or $\bar{\bar{Y}}$?

$$E(\bar{Y}) = E(\bar{\bar{Y}}) = \mu$$

$$Var(\bar{Y}) = {\sigma^2}/{3N} > Var(\bar{\bar{Y}}) = {\sigma^2}/{4N}$$

- Both are unbiased, but variance of second one is smaller ➔ estimator is more **precise**!!!

Continuous Distributions – F-distribution

- F-distribution k-samples of different sizes.
- Snedecor's F distribution is most commonly used in tests of variance (e.g., ANOVA). The ratio of two chi-squares divided by their respective degrees of freedom is said to follow an F distribution
  - k { $\{Y_{1;1}, Y_{1;2}, \ldots, Y_{1;N1}\}$ IID from a Normal($\mu_1;\sigma_1$)
  - $\{Y_{2;1}, Y_{2;2}, \ldots, Y_{2;N2}\}$ IID from a Normal($\mu_2;\sigma_2$)
  - ,...
  - $\{Y_{k;1}, Y_{k;2}, \ldots, Y_{k;N2}\}$ IID from a Normal($\mu_2;\sigma_2$)
  - $\sigma_1 = \sigma_2 = \sigma_3 = \ldots \sigma_{n_k} = \sigma$. ($1/2 \le \sigma_k/\sigma_j \le 2$)
  - Samples are independent!

Continuous Distributions – F-distribution

- F-distribution k-samples of different sizes

TABLE 10.3.2 Typical Analysis-of-Variance Table for One-Way ANOVA

Source	Sum of squares	df	Mean sum of Squares[a]	F-statistic	P-value
Between	$\sum n_i(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2$	k - 1	s_B^2	$f_0 = s_B^2/s_W^2$	pr($F \ge f_0$)
Within	$\sum (n_i - 1)s_i^2$	n_{tot} - k	s_W^2		
Total	$\sum\sum (x_{ij} - \bar{x}_{\cdot\cdot})^2$	n_{tot} - 1			

[a]Mean sum of squares = (sum of squares)/*df*

$$s_B^2 = \frac{\sum n_i(\bar{x}_i - \bar{x}_{\cdot\cdot})^2}{k-1}$$

$$s_W^2 = \frac{\sum (n_i - 1)s_i^2}{n_{tot} - k}$$

- s^2_B is a measure of variability of sample means, how far apart they are.
- s^2_W reflects the avg. internal variability within the samples.

Continuous Distributions – χ^2 [Chi-Square]

- χ^2 [Chi-Square] goodness of fit test:
  - Let $\{X_1, X_2, \ldots, X_N\}$ are IID N(0, 1)
  - $W = X_1^2 + X_2^2 + X_3^2 + \ldots + X_N^2$
  - **W ~ χ^2(df=N)**
  - Note: If $\{Y_1, Y_2, \ldots, Y_N\}$ are **IID N(μ, σ),** then

$$SD(Y) = \frac{1}{N-1}\sum_{k=1}^{N}\left(Y_k - \bar{Y}\right)^2$$

  - And the Statistics **W ~ χ^2(df=N-1)** $W = \dfrac{N-1}{\sigma^2} SD^2(Y)$

$$X^2 = \sum_{k=1}^{N}\frac{(O_k - E_k)^2}{E_k} \sim \chi 2$$

  - E(W)=N; Var(W)=2N

Continuous Distributions – Cauchy's

- Cauchy's distribution, X~Cauchy(t,s), t=location; s=scale
- PDF(X): $f(x) = \dfrac{1}{s\pi\left(1 + ((x-t)/s)^2\right)}$; $x \in \boldsymbol{R}$ (reals)
- PDF(Std Cauchy's(0,1)): $f(x) = \dfrac{1}{s\pi\left(1 + x^2\right)}$
- The Cauchy distribution is (theoretically) important as an example of a *pathological case*. Cauchy distributions look similar to a normal distribution. However, they have much heavier tails. When studying hypothesis tests that assume normality, seeing how the tests perform on data from a Cauchy distribution is a good indicator of how sensitive the tests are to heavy-tail departures from normality. The mean and standard deviation of the Cauchy distribution are undefined!!! The practical meaning of this is that collecting 1,000 data points gives no more accurate of an estimate of the mean and standard deviation than does a single point (Cauchy→T_{df}→Normal).

Continuous Distributions – Review

- Uniform, (General/Std) Normal, Student's T, F, χ^2, Cauchy distributions.
- Remained to see a good ANOVA (F-distribution Example)
- SYSTAT → File→Load (Data) →C:\Ivo.dir\Research\Data.dir\WM_GM_CSF_tissueMaps.dir\ATLAS_IVO_WM_GM.xls

  → Statistics → ANOVA →Est.Model → Dependent(Value) → Factors(Method, Hemi, TissueType)

  [For 1/2/3-Way ANOVA]

Continuous Distributions – Exponential

- Exponential distribution, X~Exponential(λ)
- The exponential model, with only one unknown parameter, is the simplest of all life distribution models.

$$f(x) = \lambda e^{-\lambda x}; \quad x \ge 0$$

- E(X)=1/ λ; Var(X)=1/ λ^2;
- Another name for the exponential mean is the **Mean Time To Fail** or **MTTF** and we have MTTF = 1/ λ.
- If *X* is the time between occurrences of rare events that happen on the average with a rate l per unit of time, then *X* is distributed exponentially with parameter λ. Thus, the exponential distribution is frequently used to model the time interval between successive random events. Examples of variables distributed in this manner would be the gap length between cars crossing an intersection, life-times of electronic devices, or arrivals of customers at the check-out counter in a grocery store.

Continuous Distributions – Exponential

- Exponential distribution, Example: ***By-hand vs. ProbCalc.htm***
- On weeknight shifts between 6 pm and 10 pm, there are an average of 5.2 calls to the UCLA medical emergency number. Let X measure the time needed for the first call on such a shift. Find the probability that the first call arrives (a) between 6:15 and 6:45 (b) before 6:30. Also find the median time needed for the first call (34.578%; 72.865%).
  - We must first determine the correct average of this exponential distribution. If we consider the time interval to be 4x60=240 minutes, then on average there is a call every 240 / 5.2 (or 46.15) minutes. Then X ~ Exp(1/46), [E(X)=46] measures the time in minutes after 6:00 pm until the first call.

Continuous Distributions – Exponential Examples

- Customers arrive at a certain store at an average of 15 per hour. What is the probability that the manager must wait at least 5 minutes for the first customer?
- The exponential distribution is often used in probability to model (remaining) lifetimes of mechanical objects for which the average lifetime is known and for which the probability distribution is assumed to decay exponentially.
- Suppose after the first 6 hours, the average remaining lifetime of batteries for a portable compact disc player is 8 hours. Find the probability that a set of batteries lasts between 12 and 16 hours.

Solutions:

- Here the average waiting time is 60/15=4 minutes. Thus X ~ exp(1/4). E(X)=4. Now we want P(X>5)=1-P(X <= 5). We obtain a right tail value of .2865. So around 28.65% of the time, the store must wait at least 5 minutes for the first customer.
- Here the remaining lifetime can be assumed to be X ~ exp(1/8). E(X)=8. For the total lifetime to be from 12 to 16, then the remaining lifetime is from 6 to 10. We find that P(6 <= X <= 10) = .1859.

Recall the example of Poisson approx to Binomial

- Suppose P(defective chip) = 0.0001=10^{-4}. Find the probability that a lot of 25,000 chips has > 2 defective!
- Y~ Binomial(25,000, 0.0001), find P(Y>2). Note that Z~Poisson(λ =n p =25,000 x 0.0001=2.5)

$$P(Z>2)=1-P(Z\le 2)=1-\sum_{z=0}^{2}\frac{2.5^z}{z!}e^{-2.5}=$$

$$1-\left(\frac{2.5^0}{0!}e^{-2.5}+\frac{2.5^1}{1!}e^{-2.5}+\frac{2.5^2}{2!}e^{-2.5}\right)=0.456$$

Normal *approximation* to Binomial

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

Normal *approximation* to Binomial – Example

- **Roulette wheel investigation:**
- **Compute P(Y>=58), where Y~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>=10 & n(1-p)=53>10 Normal approx is justified.**

Roulette has 38 slots 18red 18black 2 neutral

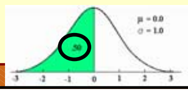
- **Z=(Y-np)/Sqrt(np(1-p)) = 58 – 100*0.47)/Sqrt(100*0.47*0.53)=2.2**
- **P(Y>=58) ➔➔ P(Z>=2.2) = 0.0139**
- **True P(Y>=58) = 0.177, using SOCR (demo!)**
- **Binomial approx useful when no access to SOCR avail.**

Normal *approximation* to Poisson

- **Let X_1~Poisson(λ) & X_2~Poisson(μ) ➔X_1+X_2~Poisson($\lambda+\mu$)**
- **Let $X_1, X_2, X_3, \ldots, X_k$ ~ Poisson(λ), and independent,**
- **$Y_k = X_1 + X_2 + \cdots + X_k$ ~ Poisson($k\lambda$), $E(Y_k)$=Var(Y_k)=$k\lambda$.**
- **The random variables in the sum on the right are independent and each has the Poisson distribution with parameter λ.**
- **By CLT the distribution of the standardized variable $(Y_k - k\lambda) / (k\lambda)^{1/2}$ ➔ N(0, 1), as k increases to infinity.**
- **So, for $k\lambda$ >= 100, $Z_k = \{(Y_k - k\lambda) / (k\lambda)^{1/2}\}$ ~ N(0,1).**
- **➔ Y_k ~ N($k\lambda$, $(k\lambda)^{1/2}$).**

Normal approximation to Poisson – example

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



Poisson or Normal approximation to Binomial?

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

$$\binom{n}{y} p_n^{\ y} (1-p_n)^{n-y} \xrightarrow[n \times p_n \longrightarrow \lambda]{\text{WHY?} \quad n \longrightarrow \infty} \frac{\lambda^y e^{-\lambda}}{y!}$$

  - 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

Exponential *family* and arrival numbers/times

- First, let T_k denote the time of the k'th arrival for k = 1, 2, ... The *gamma experiment* is to run the process until the k'th arrival occurs and note the time of this arrival.
- Next, let N_t denote the number of arrivals in the time interval $(0, t]$ for $t \geq 0$. The *Poisson experiment* is to run the process until time t and note the number of arrivals.
- How are T_k & N_t related?
- $N_t \geq k$ ← → $T_k \leq t$

density function of the k'th arrival time is $f_k(t) = (rt)^{k-1} r e^{-rt} / (k-1)!$, $t > 0$. This distribution is the gamma distribution with shape parameter k and rate parameter r. Again, 1/r is knows as the scale parameter. A more general version of the gamma distribution, allowing non-integer k,

Independence of continuous RVs

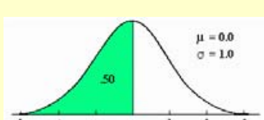
- The RV's $\{Y_1, Y_2, Y_3, \ldots, Y_n\}$ are independent if for any n-tuple $\{y_1, y_2, y_3, \ldots, y_n\}$

$$P(\{Y_1 \leq y_1\} \cap \{Y_2 \leq y_2\} \cap \{Y_3 \leq y_3\} \cap \ldots \cap \{Y_n \leq y_n\}) =$$
$$= P(Y_1 \leq y_1) \times P(Y_2 \leq y_2) \times P(Y_3 \leq y_3) \times \ldots \times P(Y_n \leq y_n)$$

Standard Normal Curve

- The standard normal curve is described by the equation:

$$y = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$



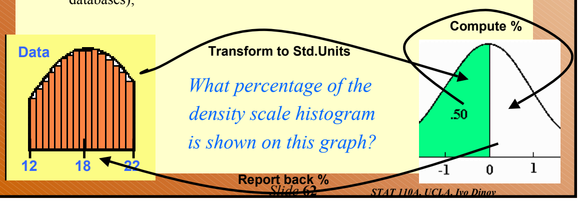
Where remember, the natural number e ~ 2.7182…

We say: X~Normal(μ, σ), or simply X~N(μ, σ)

Quincunx – Galton Board

Standard Normal Approximation

- The standard normal curve can be used to estimate the percentage of entries in an interval for any process. Here is the protocol for this approximation:
  - Convert the interval (we need the assess the percentage of entries in) to standard units. We saw the algorithm already.
  - Find the corresponding area under the normal curve (from tables or online databases);

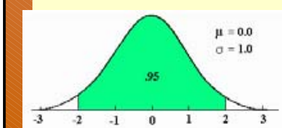


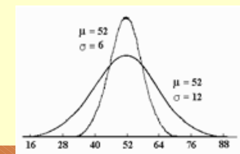
General Normal Curve

- The general normal curve is defined by:
  - Where μ is the average of (the symmetric) normal curve, and σ is the standard deviation (spread of the distribution).

$$y = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$$

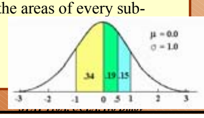
  - Why worry about a standard and general normal curves?
  - How to convert between the two curves?



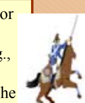


Areas under Standard Normal Curve – Normal Approximation

- Protocol:
  - Convert the interval (we need to assess the percentage of entries in) to Standard units. Actually convert the end points in Standard units.
    - In general, the transformation $X \rightarrow (X-\mu)/\sigma$, standardizes the observed value X, where μ and σ are the average and the standard deviation of the distribution X is drawn from.
  - Find the corresponding area under the normal curve (from tables or online databases);
    - Sketch the normal curve and shade the area of interest
    - Separate your area into individually computable sections
    - Check the Normal Table and extract the areas of every sub-section
    - Add/compute the areas of all sub-sections to get the total area.



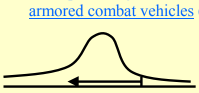
Areas under Standard Normal Curve – Example



- Many histograms are similar in shape to the standard normal curve. For example, persons height. The height of all incoming female army recruits is measured for custom training and assignment purposes (e.g., very tall people are inappropriate for constricted space positions, and very short people may be disadvantages in certain other situations). The mean height is computed to be 64 *in* and the standard deviation is 2 *in*. Only recruits shorter than 65.5 in will be trained for tank operation and recruits within ½ standard deviations of the mean will have no restrictions on duties.
  - What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?
  - About what percentage of the recruits will have no restrictions on training/duties?

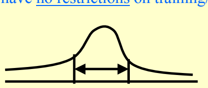
Areas under Standard Normal Curve - Example

- The mean height is 64 *in* and the standard deviation is 2 *in*.
  - Only recruits shorter than 65.5 in will be trained for tank operation. What percentage of the incoming recruits will be trained to operate armored combat vehicles (tanks)?



X → (X-64)/2
65.5 → (65.5-64)/2 = ¾
Percentage is 77.34%

  - Recruits within ½ standard deviations of the mean will have no restrictions on duties. About what percentage of the recruits will have no restrictions on training/duties?



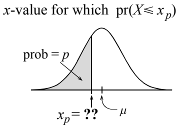
X → (X-64)/2
65 → (65-64)/2 = ½
63 → (63-64)/2 = -½
Percentage is 38.30%

The inverse problem – Percentiles/quantiles

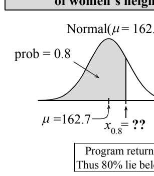
(a) *p*-Quantile

Programs supply x_p

x-value for which $pr(X \leq x_p) = p$



(b) 80th percentile (0.8-quantile) **of women's heights**



80% of people have height below the 80th percentile. This is EQ to saying there's 80% chance that a random observation from the distribution will fall below the 80th percentile.

(c) Further percentiles of women's heights

Percent	1%	5%	10%	20%	30%	70%	80%	90%	95%
Propn	0.01	0.05	0.1	0.2	0.3	0.7	0.8	0.9	0.95
Percentile	*(or quantile)*								
(cm)	148.3	152.5	154.8	157.5	159.4	166.0	167.9	170.6	172.9
(ft'in")	4'10"	5'0"	5'0"	5'2"	5'2"	5'5"	5'6"	5'7"	5'8"

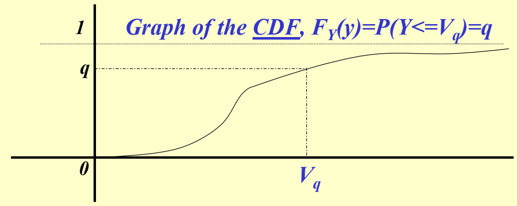
The inverse problem is what is the height for the 80th percentile/quantile? So far we studied given the height value what's the corresponding percentile?

Identifying Common Distributions – QQ plots

- **Plots** are useful for identifying candidate distribution model(s) in approximating a population (data) distribution.
- **Histograms**, can reveal much of the features of the data distribution.
- **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.
- E.g., $q=0.80$, $Y=\{1,2,3,4,5,6,7,8,9,10\}$. The q-th quantile $V_q= 8$, since 80% of the data is at or below 8.

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.



Constructing QQ plots

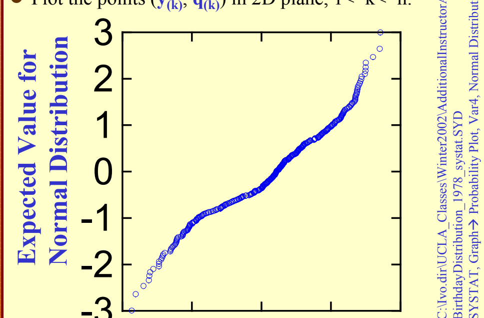
- **Start off with data** $\{y_1, y_2, y_3, \ldots, y_n\}$
- **Order statistics** $y_{(1)} <= y_{(2)} <= y_{(3)} <= \ldots <= y_{(n)}$
- Compute **quantile rank**, $q_{(k)}$, for each observation, $y_{(k)}$,

$$P(Y<= q_{(k)}) = (k-0.375) / (n+0.250),$$

  where Y is a RV from the (target) model distribution.
- Finally, plot the points $(y_{(k)}, q_{(k)})$ in 2D plane, $1<=k<=n$.
- **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:
  $P((model)Y<=(data)y_{(1)}) \sim 1/n$;
  $P(Y<=y_{(2)}) \sim 2/n$; $P(Y<=y_{(3)}) \sim 3/n$; …

Example - Constructing QQ plots

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

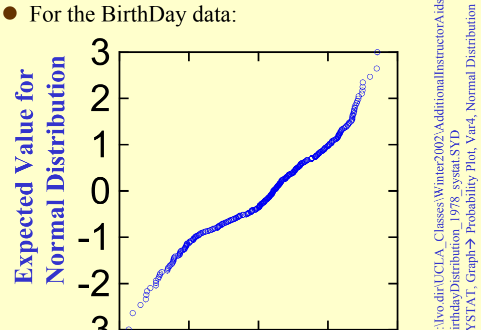


Data transformations

- In practice oftentimes observed **data does not directly fit any of the models** we have available. In these cases transforming the raw data may provide/satisfy the requirements for using the distribution models we know.
- Common transformations: Y=T(X), X=raw data, Y=new
  - Data positively skewed to right use T(X)=Sqrt(X) or T(X)=log(X)
  - If data varies by more than 2 orders of magnitude
    - For X>0, use T(X)=log(X)
    - For any X, use T(X)= –1/X.
    - If X are counts (categorical var's), T(X)=Sqrt(X)
    - X=proportions & largest/ smallest Proportions >=2, use Logit transform: T(X) = log[X/(1-X)].

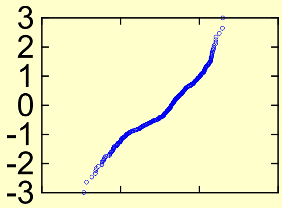
Data transformations - Example

- For the BirthDay data:



Data transformations - Example

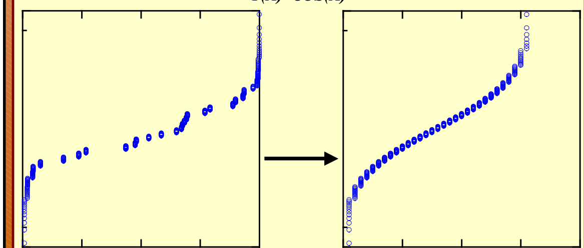
- BirthDay data: C:\Ivo.dir\UCLA_Classes\Winter2002\AdditionalInstructorAids BirthdayDistribution_1978_systat.SYD SYSTAT, Graph→ Probability Plot, Var4, Normal Distribution



Data transformations - Example

- BirthDay data: C:\Ivo.dir\UCLA_Classes\Winter2002\AdditionalInstructorAids BirthdayDistribution_1978_systat.SYD SYSTAT, Graph→ Probability Plot, COS(Var2), Normal Distribution

T(X)=COS(X)



Student's *t*-distribution

- For random samples from a Normal distribution,

$$T = \frac{(\overline{X} - \mu)}{SE(\overline{X})}$$

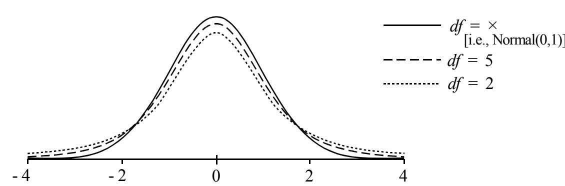
Recall that for samples from N(μ , σ)

$$Z = \frac{(\overline{X} - \mu)}{SD(\overline{X})} = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} \sim N(0,1)$$

is exactly distributed as Student($df = n - 1$) ← Approx/Exact Distributions

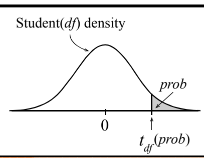
- but methods we shall base upon this distribution for T work well even for small samples sampled from distributions which are quite non-Normal.
- df is number of observations –1, degrees of freedom.

Density curves for Student's *t*



Student(df) density curves for various *df*.

Reading Student's *t* table



Desired *upper-tail prob*

TABLE 7.6.1 Extracts from the Student's t-Distribution Table

prob

df	.20	.15	.10	.05	.025	.01	.005	.001	.0005	.0001
6	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959	8.025
7	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7.063
8	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041	6.442
...	...	...	...	...	...	...	...	...	...	...
10	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587	5.694
...	...	...	...	...	...	...	...	...	...	...
15	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073	4.880
...	...	...	...	...	...	...	...	...	...	...
∞	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.719

Desired *df* →

t-value

Do we need an simulation of ***T*** and ***Z*** scores? Use the Online compute-engine …

Practice Problems – Areas under Normal Curve

Ex 1) Z = a standard normal R.V.

(a) $P(Z < 1.43) =$

(b) $P(Z > -0.89) =$

(c) $P(-2.16 < Z < -0.65) =$

Ex 2) $X \sim$ normal, $\mu = 30$, $\sigma = 6$

(a) $P(X > 17) =$

(b) $P(X < 22) =$

(c) $P(32 < X < 41) =$

Application of Normal Distribution

Ex 3)

Soft-drink machine; μ= 200(milliliters/cup), σ = 15

(a) P(a cup will contain more than 224)

(b) P(a cup contains between 191 and 209)

(c) How many cups will overflow if 230 milliliter cups used for the next 1000 drinks?

(d) Below what value do we get the smallest 25% of the drinks?

Normal Approximation to Binomial

- Poisson vs. Binomial
- Normal vs. Binomial
- Theorem X= binomial, $\mu = np$, $\sigma^2 = npq$

  Then, $Z = \dfrac{X-np}{\sqrt{npq}} \sim n(z;0,1)$ as $n \to \infty$

$$P\{X=k\} = P\{k-0.5 < X < k+0.5\} \approx P\left\{\frac{k-0.5-\mu}{\sigma} < Z < \frac{k+0.5-\mu}{\sigma}\right\}$$

$$P\{k_1 \le X \le k_2\} \approx P\{\frac{k_1-0.5-\mu}{\sigma} \le Z \le \frac{k_2+0.5-\mu}{\sigma}\}$$

- Approximation becomes better as n gets larger $p \to 0.5$

Normal Approximation to Binomial

Ex 1) $X \sim$ binomial, $n = 15$, $p = 0.4$ Find $P(X=7)$

(1) binomial

(2) normal

Ex 2)

$X \sim$ binomial, $n = 15$, $p = 0.2$

→ Find $P(1 \le X \le 4)$

Gamma and Exponential Distributions

- **Gamma Distribution**
  - **Gamma function :** $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \quad \alpha > 0$

Properties

$$\begin{cases} \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1) \\ \Gamma(n) = (n-1)! \quad \text{for positive integer } n \\ \Gamma(1) = 1 \quad\quad \Gamma(0.5) = \sqrt{\pi} \end{cases}$$

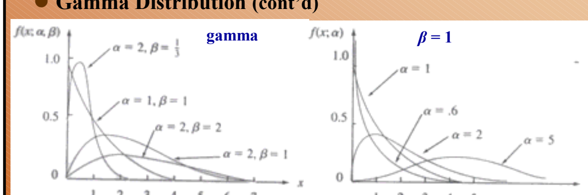
\- $X \sim$ Gamma with parameters α and β if

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

where $\alpha > 0$, $\beta > 0$

Gamma and Exponential Distributions

- **Gamma Distribution (cont'd)**



- $E(X) = \alpha\beta \quad Var(X) = \alpha\beta^2$
- If $\alpha = 1$; $\quad f(x) = \dfrac{1}{\beta} e^{-\frac{x}{\beta}}, \quad x > 0, \ \beta > 0$
- Cdf: incomplete gamma function

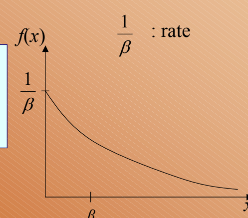
Ex 4) $X \sim$ Gamma, $\alpha = 2, \beta = 1$ → Find $P(1.8 < X < 2.4)$

Gamma and Exponential Distributions

- **Exponential Distribution**
  - Useful in modeling time between arrivals at service facilities
  - One Parameter ; β
  - a special case of Gamma

$\dfrac{1}{\beta}$: rate

$$- f(x) = \begin{cases} \dfrac{1}{\beta} e^{-\frac{1}{\beta}x}, & x > 0 \\ 0 & , \text{otherwise} \end{cases}$$



- $E(X) = \beta \quad Var(X) = \beta^2$

  mean=standard deviation

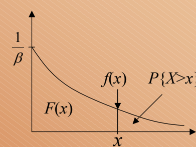
Gamma and Exponential Distributions

● Exponential Distribution (cont'd)

- CDF : $F(x) = P(X \le x) = \int_0^x \frac{1}{\beta} e^{-\frac{x}{\beta}} dx = 1 - e^{-\frac{x}{\beta}}, \; x > 0$

- Tail probability

$P(X > x) = 1 - F(x) = e^{-\frac{x}{\beta}}, \quad x > 0$



Ex 1) X = response time at a certain on-line computer terminal

$X \sim$ exponential with $E(X) = 5$(sec.).

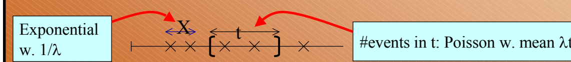
(a) $P(X \le 10) =$

(b) $P(5 \le X \le 10) =$

Gamma and Exponential Distributions

Relationship to the Poisson Process

# of events in any time interval t has a Poisson distribution w/ parameter λt → the distribution of **the elapsed time between two successive events** is exponential with parameter $\beta = \frac{1}{\lambda}$



Why? Poisson : P(no events in t) = $P(0;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$

Let X = time until the first event. $P(X > t) = e^{-\lambda t}$

Then P(no events in t) =

i.e., $P(0 \le X \le t) = 1 - e^{-\lambda t}$ = CDF of exponential with $\lambda = \frac{1}{\beta}$ or $\beta = \frac{1}{\lambda}$

Gamma and Exponential Distributions

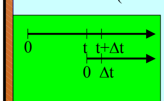
●Model for Component Lifetime

■Exponential Dist' is useful due to "**Memoryless property**"

■Memoryless property

if T~ exponential with ($\beta > 0$)

$P(T > t + \Delta t \mid T > t) = \frac{P(T > t + \Delta t, T > t)}{P(T > t)} = \frac{P(T > t + \Delta t)}{P(T > t)}$



$= \frac{e^{-\frac{t+\Delta t}{\beta}}}{e^{-\frac{t}{\beta}}} = e^{-\frac{\Delta t}{\beta}} = P(T > \Delta t)$

- The distribution of the additional lifetime = The original distribution of lifetime (Memoryless Property)

Gamma and Exponential Distributions

Ex 2) Hotline

calls received at a hotline ~ Poisson with = 0.5/day.

X = # days between successive calls.

(a) $P(X > 2) =$

(b) $P(X > 5 \mid X > 3) =$

Ex 3)

T (= time to failure (in years) of a component)

~ exponential with $\beta = 5$

(a) $P(T > 8) =$

(b) 5 components are installed.

P(at least 2 are functioning at the end of 8 years) =

Chi-Square Distribution

● **Special case of Gamma**

■ $\alpha = n/2, \beta = 2$ where n = a positive integer

■ $X \sim$ Chi-Square with parameter n (degree of freedom) if

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & , \; x > 0 \\ 0 & , \; \text{otherwise} \end{cases}$$

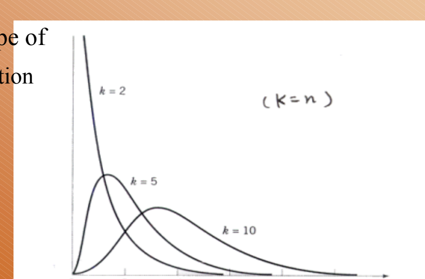
● $E(X) = \alpha\beta = \frac{n}{2} \cdot 2 = n$

$Var(X) = \alpha\beta^2 = \frac{n}{2} \cdot 2^2 = 2n$

$X = \sum_{i=1}^{n} Z_i^2$ follows χ^2 with df=n, where Z_i are iid N(0,1)

Chi-Square Distribution

- Useful in statistical inference, hypothesis testing
- The shape of χ^2 Distribution



Probability density functions of several χ^2 distributions.

Lognormal Distribution

- $X \sim$ lognormal with parameters μ and σ, if

$$\ln(X) \sim N(x; \mu, \sigma)$$

- $$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\,\sigma\, x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & x \ge 0 \\ 0 & \text{, otherwise} \end{cases}$$

- $E(X) = \exp(\mu + \sigma^2/2)$

  $Var(X) = \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}$

Ex) Let X ~ lognormal with parameter $\mu = 3.2$ and $\sigma, = 1$

$P(X > 8) =$

Weibull Distribution

- $X \sim$ Weibull Distribution with parameters α and β if

$$f(x) = \begin{cases} \alpha\beta\, x^{\beta-1} e^{-\alpha x^{\beta}} & , x > 0 \\ 0 & , \text{otherwise} \end{cases}$$

- If $\beta = 1$; $f(x) = \alpha e^{-\alpha x}$ (exponential with parameter $\frac{1}{\alpha}$)

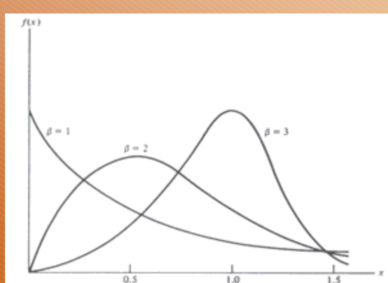
$$E(X) = \alpha^{-\frac{1}{\beta}} \Gamma(1 + \frac{1}{\beta}) \qquad F(x) = 1 - e^{-\alpha x^{\beta}}$$

$$Var(X) = \alpha^{-\frac{2}{\beta}} \{\Gamma(1 + \frac{2}{\beta}) - [\Gamma(1 + \frac{1}{\beta})]^2\}$$

- Useful in Reliability, life testing problems

Weibull Distribution

- The shape of Weibull Distribution



Weibull distributions ($\alpha = 1$).

Weibull Distribution

Ex)

X = service life of a battery ~ Weibull, $\alpha = \frac{1}{2}$, $\beta = 2$

(a) Expected service life?

(b) P(a battery will still be operating after 2 years) = ?

Beta Distribution

- Provides positive density only in an interval of finite length

$X \sim$ Beta Distribution with parameters α and β if

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & , 0 < x < 1 \ (\alpha > 0,\ \beta > 0) \\ 0 & , \text{otherwise} \end{cases}$$

$$E(X) = \frac{\alpha}{\alpha + \beta}, \quad Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Ex)

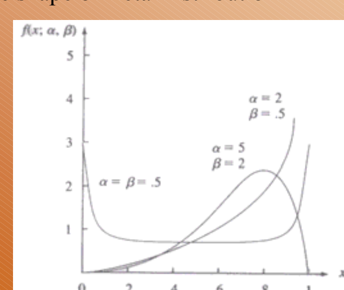
X = proportion of TV sets requiring service during the first year

~ beta, $\alpha = 3$, $\beta = 2$.

P(at least 80% of the model sold this year will require service in 1 year)

Beta Distribution

- The shape of Beta Distribution



Relation among Distributions