Introduction to the Theory of Statistics Part 2

PM522b

Meredith Franklin

Division of Biostatistics, University of Southern California

Slides 1, 2020

Course Details

- Book: Statistical Inference, 2nd Ed. Casella G and Berger RL. Wadsworth & Brooks, 2002
- Lecture slides will be posted on Blackboard
- Additional handouts will be posted as needed
- We will cover CB Chapter 5 properties of random samples, order statistics, and Chapters 6-12
- More on the theory of regression than presented in CB (hopefully!)
- We will do some computation and visualization using R
- Grading: Homework (Weekly, 30%), Midterm Exam (30%), Final Exam (35%), In-Class Participation (5%)

Course Details

- Software: we will use R
  - functions for distributions
  - writing custom functions
  - sampling data
  - simulating data
  - estimation
- Homework will mostly be handwritten solutions, but some computation.
- Exams will be in-class. You can bring one "cheat sheet" of paper to the exams (8.5x11, double sided).

Topics Covered in these Slides

1. Introduction to statistics and statistical inference
    - Review of random variables, cdf, pmf, pdf
    - Bridging from probability to inference
2. Review of random samples, functions of random variables
    - Relating samples to populations
    - Empirical distribution functions
    - Graphical representations of statistics
    - Order statistics
    - Sampling from the normal distribution and the derived distributions (t, χ^2, F)

Random Variables and Distribution Functions (CB Sections 1.4-1.6, 2.4)

Random Variables

- Probabilities describe the population – probability mass/density functions are a way to mathematically characterize the population.
- Statistical inference is the process of characterizing populations using data. In this course, we'll assume that our sample is a random draw from the population.

A **random variable** is a numerical outcome of an experiment, and comes in two varieties **discrete** and **continuous**.
Discrete random variables take on only a countable number of possibilities. Mass functions will assign probabilities that they take specific values.
Continuous random variable can conceptually take any value on the real line or some subset of the real line and we talk about the probability that they lie within some range. Density functions characterize these probabilities.

Random Variables

Examples of discrete random variables:

- Coin toss (binomial)
- Number of hospitals in the neighbourhoods of LA county (count, Poisson)

Examples of continuous random variables:

- Height, weight, temperature (normal, lognormal)
- Time to event (exponential, Weibull)

Recall we use upper case X to denote a random, unrealized version of the random variable and a lowercase x to denote a specific realization or number that we plug in.

For all random variables we need mathematical functions to model the probabilities of collections of realizations. We use probability mass and density functions and take possible values of the random variable and assign the associated probabilities.

CDF

The cumulative distribution function (cdf) for a discrete random variable evaluated at x is **the probability that a random variable X will take a value less than or equal to x**.

$$F_X(x) = P(X \leq x), \forall x$$

which has three conditions:

1. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$
2. $F_X(x)$ is a non-decreasing function of x
3. $F_X(x)$ is right continuous

For continuous random variables, $F_X(x)$ is a continuous function of X and the CDF is represented as $F_X(x) = \int_{-\infty}^{x} f(t)dt$.

We can say a random variable X is continuous if $F_X(x)$ is a continuous function of x. Similarly a random variable X is discrete if $F_X(x)$ is a step function of x. Note, the survival function of a random variable X is defined as the probability that the random variable is greater than the value x: $S_X(x) = P(X > x)$. Since the survival function evaluated at a particular value of x is calculating the probability of the opposite event, we have $S_X(x) = 1 - F_X(x)$.

PMF

A probability mass function (pmf) is the primary means of describing a discrete probability distribution. The pmf evaluated at a value corresponds to the probability that a random variable takes that value.

- The pmf of a discrete random variable X:

$$f_X(x) = P(X = x), \forall x$$

  To be a valid pmf, the probability must satisfy:

  1. $f_X(x) \geq 0 \forall x$ (all probabilities must be positive)
  2. $\sum_x f_X(x) = 1$ (the sum is taken over all values of x)
  3. $P(X \in A) = \sum_{x \in A} f_X(x)$ (to determine the probability of event A, sum up the probabilities of the x values in A).

Example

X is the result of flipping a coin where X=0 is tails and X=1 is heads. If the coin is fair, $P(x) = (1/2)^x (1/2)^{1-x}$ for $x = 0, 1$
If we do not know whether the coin is fair or not, $P(x) = \theta^x (1-\theta)^{1-x}$ for $x = 0, 1$

PDF

A probability density function (pdf) is a function associated with a continuous random variable. Areas under pdfs correspond to probabilities for a random variable.

- The pdf of a continuous random variable X is the function that satisfies:

$$F_X(x) = \int_{-\infty}^{x} f(t)dt$$

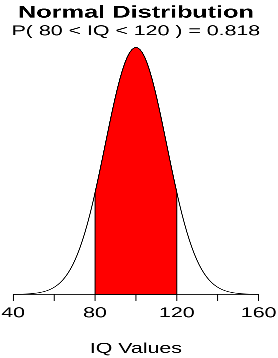
  And hence,

$$\frac{dF_X(x)}{dx} = f(x)$$

- Using the fundamental theorem of calculus, the derivative of the cdf is the pdf (when f(x) is continuous).
- To be a valid pdf, the function f must satisfy
  1. $f_X(x) \geq 0 \forall x$ (must be greater than or equal to zero everywhere)
  2. The total area under $f_X(x)$ is equal to one

PDF

Example: Children's IQ scores are normally distributed with mean 100 and standard deviation of 15. We can find the probability of having an IQ between, less than, or greater than some values by calculating the area under the curve. Probability of having IQ between 80 and 100 is 0.818%

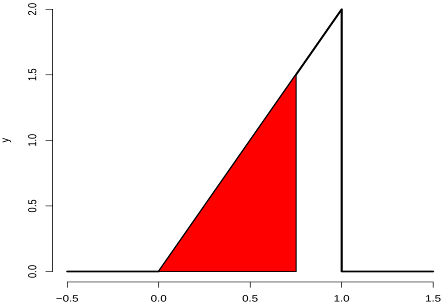


Recall, for a continuous distribution, the probability of an exact value of x is 0.

PDF

We know how to calculate the area under a normal distribution by converting to the standard normal distribution, getting z-values, and looking up values in tables. Let's look at a distribution where we actually calculate the area under the curve to generate probabilities:

$$f_X(x) = \begin{cases} 2x, 0 < x < 1 \\ 0, \text{otherwise} \end{cases}$$



PDF

This is a mathematically valid density because it is non-negative (y is above 0 everywhere), and the area under the density is the area of a triangle $0.5 * \text{base} * \text{height} = 0.5 * 1 * 2 = 1$.
We can use this to calculate probabilities, such as what is the probability that x is less than 0.75? If we want to put this into context of a real example, suppose X is the time in minutes to take a 1 hour exam. So the probability of finishing the exam in three quarters of an hour (0.75) is $0.5 * 0.75 * 1.5 = 0.5625$.
Relation to CDF:
Recall the equation for the CDF

$$F_X(x) = P(X \leq x), \forall x$$

The probability that $X < 0.75$ is similarly calculated as we did on the previous slide for the PDF, $0.5 * \text{base} * \text{height} = 0.5 * 0.75 * 1.5 = 0.5625$

Distribution functions in R

A list of the distribution functions in R, along with how to generate cdf, pmf, and pdf and some examples are available here:
`http://meredithfranklin.github.io/R-Probability-Distributions.html`

Expectation

The expected value (CB 2.2), which is the (population) mean, is the center of a distribution.

- for discrete random variables $E[X] = \sum_x x f_X(x)$, where the sum is taken over the possible values of x.
- for continuous random variables $E[X] = \int\limits_{-\infty}^{\infty} f(x)dx$.

Often we use the parameter μ to represent the population mean.

Variance

The variance, is a measure of the spread of a distribution around its mean.

- for random variable X with mean μ, the variance of X is $\text{Var}(X) = E[(X-\mu)^2] = E[X^2] - E[X]^2$

Often we use the parameter σ^2 to represent the population variance. The positive square root of the variance is the standard deviation. It is often easier to interpret because it has the same units as the random variable.
Small values of the variance mean that X is very likely to be close to $E[X]$.

Be sure to review the properties of the mean and variance (CB 2.2, 2.3).

Linking Distributions and Random Samples (CB 5.1, 5.2)

Distributions and parameters

- In PM522a you learned specific discrete (Discrete Uniform, Hypergeometric, Binomial, Poisson, Negative Binomial, Geometric) and continuous (Uniform, Gamma, Exponential, Normal, Beta, Lognormal) distribution functions.
- The parameters of these functions were assumed to be known.
- Using a pdf with known parameters, we can say something about a random variable X

Example

$X \sim f_X(x|\theta), x \in R$ and $\theta \in \Theta$ are parameters
If $f_X(x|\theta)$ is the binomial distribution then we know $X \sim \text{binomial}(n, p)$ where n and p are our parameters
$\theta = (n, p)$
Furthermore we can calculate $E(X) = np$ and $V(X) = np(1-p)$

Distributions and Random Samples

Knowing the value of the parameter, we can evaluate the distribution function.

A Numerical Example

What is the probability that a family of 3 children will have 2 girls given that the probability of having a girl is $1/2$?
In R: choose(3, 2) * 0.5^2*0.5^1 OR dbinom(2,3,1/2) $=$0.375

A Less Obvious Example

Suppose we toss a coin 10 times and observe 8 heads. What is the probability of heads?
If the coin was perfectly fair, then we could assume $\theta = 1/2$. But a) we don't know anything about the coin, and b) having flipped $8/10$ heads does not support that P(heads)$=1/2$.

Distributions and Random Samples

- The two previous examples illustrate a sequence of n Bernoulli trials.
- The distribution for n independent Bernoulli trials is Binomial(p), meaning there is only one parameter $\theta = p$ in the distribution because we known n.
- The normal distribution has two parameters, $\theta = (\mu, \sigma^2)$.

Example

$X \sim f_X(x|\theta), x \in R$ and $\theta \in \Theta$ are parameters
If $f_X(x|\theta)$ is the normal distribution then we know $X \sim N(\mu, \sigma^2)$ where μ and σ are our parameters
$\theta = (\mu, \sigma^2)$
Furthermore we can find properties of these parameters $E(X) = \mu$ and $V(X) = \sigma^2$

Distributions and Random Samples

- We need to bridge from probability to (inferential) statistics.
- Populations to samples: data.
- Experiments are performed to collect information (data) from which we can (imperfectly) understand the population.
- A random sample is drawn from our population and we need a suitable function to describe the population from the sample.
- We want to make *inference* about a population based on information contained in this random sample.
- Always remember: the sample is NOT the population.

Random Samples

- In statistics and statistical inference, we have random samples of X.
- We don't know the pdf of X but want to be able to say something about its distribution.
- A random variable could be represented by any possible pdf, however one model will be more probable than the others.
- $\mathbf{X} = (X_1, X_2, ..., X_n)$ is a set of iid random variables with an unknown distribution function.
- $X \sim f_X(x|\theta), x \in R$ and $\theta \in \Theta$ and we further define $\Theta \in R^d$ as the parameter space.
- We regard $f_X(x|\theta)$ as the parametric model function.

Random Samples

- The objective of statistical inference is thus to assess aspects of our unknown parameters $\boldsymbol{\theta}$ given random samples.
- Notation: X and X_i represent random variables; x and x_i represent observed values of the random variable X.
- Notation: boldface denotes multiple variates where $\mathbf{X}$ represents random variables $(X_1, X_2, ..., X_n)$, and $\mathbf{x}$ represents observations $(x_1, x_2, ..., x_n)$
- There are three major components to statistical inference: point estimation, confidence/interval estimation, and hypothesis testing.
- Point estimation is a single value estimate of θ_i computed from the data x.
- Confidence estimation provides a set of values having a probability of including the true (but unknown) value of θ_i.
- Hypothesis testing involves setting up a hypothesis about θ_i and assessing the plausibility of the hypothesis using the data x.
- We will also focus on the theory of linear regression and ANOVA at the end of the semester.

Frequentist vs Bayesian Inference

Two types of inference exist: Frequentist and Bayesian
In the context of understanding the unknown parameter θ given random samples, we can describe the two approaches. Suppose the unknown parameter of interest is the mean μ of a normal distribution and we have observations $x_1, x_2, ...x_n$:

- Frequentist approach:
  - We do not make any further probabilistic assumptions on the parameter
  - Treat μ as a fixed but unknown constant
  - Use data reduction techniques to summarize the information in the sample (i.e. sample mean). This summary is a function which is also known as a statistic.
  - The data are a repeatable random sample. That is, sampling is infinite.
  - Assessment of the suitability of the estimate for our unknown parameter is based in how it would perform if done repeatedly (frequency interpretation)

Frequentist vs Bayesian Inference

Two types of inference exist: Frequentist and Bayesian
In the context of understanding the unknown parameter θ given random samples, we can describe the two approaches. Suppose the unknown parameter of interest is the mean μ of a normal distribution and we have observations $x_1, x_2, ...x_n$:

- Bayesian approach:
  - Treat μ as having a probability distribution, not fixed.
  - The prior distribution on the unknown parameter is either known, assumed on some information, or drawn from thin air.
  - The uncertainty in μ is taken into account with the prior, without using the observations.
  - Use Bayes' theorem to modify the probability of our unknown parameter given the observations.
  - The posterior distribution is the modified prior distribution of the unknown μ.

Random Variables, Functions, and Samples

- The classical, frequentist approach is concerned with experiments that are replicated a fixed number of times.
- Replication means that each repetition is performed under identical conditions and is mutually independent (iid).
- We use the sample to extract information used to draw inferences about the population.

Empirical Distribution Function

- For discrete probability distributions we can define the empirical distribution function (edf).

Empirical Distribution Function (edf)

Let our sample $x_1, x_2, ...x_n$ be iid random variables with cdf F_n
The edf associated with the sample $\hat{F}_n$ is the discrete distribution function defined by assigning probability $1/n$ to each x_i
Example edf: A fair die is rolled $n = 20$ times resulting in the sample $x = 1, 2, 3, 6, 3, 4, 5, 2, 5, 1, 2, 4, 4, 2, 3, 5, 6, 1, 2, 6$ the edf $\hat{P}_{20}$ assigns the probabilities:

x_i	$\#x_i$	$\hat{P}_{20}(x_i)$
1	3	0.15
2	5	0.25
3	3	0.15
4	3	0.15
5	3	0.15
6	3	0.15

Empirical Distribution Function

- The true probabilities are $1/6$ but the empirical probabilities range from 0.15 to 0.25
- The fact that the empirical probabilities $\hat{P}_n$ differ from P_n is sampling variation
- $\hat{P}_n(A) = \#\{x_i \in A\}\frac{1}{n}$
- The empirical cumulative distribution function associated with $\hat{P}_n$ is denoted $\hat{F}_n$

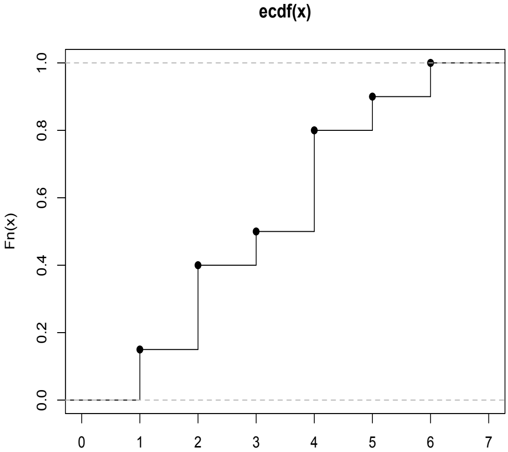
Definition: Empirical cdf

$$\hat{F}_n(a) = \hat{P}_n(X \leq a) = \frac{\#\{x_i \leq a\}}{n}$$

Empirical CDF

In R:

```
x<-c(1,2,3,6,3,4,5,2,5,1,2,4,4,2,3,5,6,1,2,6)
plot.ecdf(x, verticals=TRUE)
```



Relating samples to populations: Sample Mean

- Expected values are another common estimate of the population from our random sample
- Let $E(X_i) = \mu$ denote the population mean
- We can use the plug-in principle to estimate the mean
- For our sample $x_1, x_2, ...x_n$, $\hat{\mu}_n = \sum_{i=1}^{n} \frac{x_i}{n}$

Example: mean of the empirical distribution

A fair die is rolled $n = 20$ times resulting in the sample
$x = \{1, 2, 3, 6, 3, 4, 5, 2, 5, 1, 2, 4, 4, 2, 3, 5, 6, 1, 2, 6\}$ the population mean is:
$\mu = E(X_i) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$
But the sample mean is 3.35
$\hat{\mu}_{20} \neq \mu$

Relating samples to populations: Sample Variance

- Variance is another common estimate of the population from our random sample
- Let $V(X_i) = \sigma^2$ denote the population variance
- We can use the plug-in principle to estimate the variance of the empirical distribution
- For our sample $x_1, x_2, ...x_n$, $\hat{\sigma}_n^2 = \sum_{i=1}^{n} \frac{(x_i - \hat{\mu}_n)^2}{n}$

Example: variance of the empirical distribution

A fair die is rolled $n = 20$ times resulting in the sample
$x = \{1, 2, 3, 6, 3, 4, 5, 2, 5, 1, 2, 4, 4, 2, 3, 5, 6, 1, 2, 6\}$ the population variance is:
$\sigma^2 = E(X_i^2) - (E(X_i))^2 = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - 3.5^2 = 2.92$
But the sample variance is 1.73
$\hat{\sigma}_{20}^2 \neq \sigma^2$

Relating samples to populations: Quantiles

- Quantiles are another common estimate of the population from our random sample
- The estimate of the population quantile is the corresponding quantile of the empirical distribution (e.g. median (2nd quantile or 50%) and interquartile range (3rd-1st quantile or 75%-25%))
- We can use the plug-in principle to estimate the quantiles of the empirical distribution

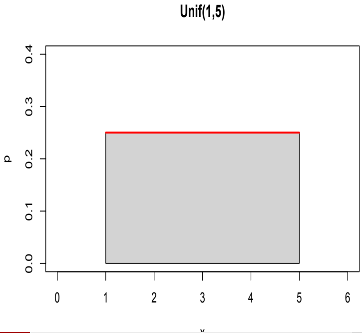
Example: quantiles of the empirical distribution

If we take $n = 20$ draws from a Uniform distribution $X \sim U(1, 5)$ resulting in the sample x=$\{$4.92, 4.89, 1.93, 2.25, 3.08, 2.58, 3.91, 3.11, 2.56, 1.16, 3.55, 3.57, 1.16, 1.02, 2.20, 4.80, 4.94, 4.99, 2.68, 4.58$\}$ the population quantiles are:
$Pr[X \leq x] \geq q$ and $Pr[X \geq x] \geq 1 - q$ where q is the qth quantile, $0 < q < 1$
For a continuous r.v., $F(x) = q$, so for $X \sim U(1, 5), F(x) = 1/2$ when $x = 3$

Relating samples to populations: Quantiles

In R: Uniform distribution on interval [1,5]

```
x=seq(1,5,length=200)
y=rep(1/(5-1),200) # f(x)=1/(b-a)
plot(x,y,type="l",xlim=c(0,6),ylim=c(0,0.4),
lwd=2,col="red",ylab="p",main="Unif(1,5)")
polygon(c(1,x,5),c(0,y,0),col="lightgray")
lines(x,y,type="l",lwd=2,col="red")
```



Relating samples to populations: Quantiles

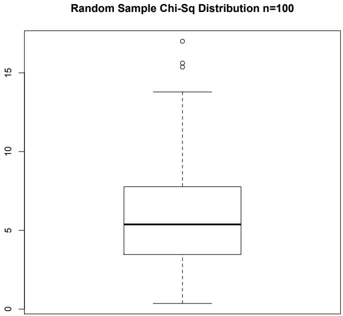
The q^{th} quantile of a distribution with distribution function F is the point x_q so that

$$F(x_q) = q$$

So the 0.75 quantile of a distribution is the point such that 75% of the mass of the density lies below it. In other words, it is the point where the probability of getting a randomly sampled point below it is 0.75.
In R we can get quantiles for the common distributions with the prefix q in front of the distribution name. For example, the 0.75 quantile of $U(1, 5)$ is found using `qunif(0.75,1,5)`. The answer is 4.

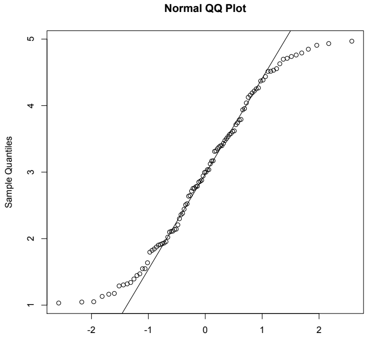
Graphical Representations

- Graphical uses of quantiles can be useful in determining aspects of the population from our random sample
- Box plots: gives an indication of symmetry of distribution
  - create a box around the 1st and 3rd quartile (25% and 75%)
  - add a line at the median (50%)
  - extend whiskers to extreme values (1.5 iqr or 5%-95%)
  - add outliers as points beyond the whiskers

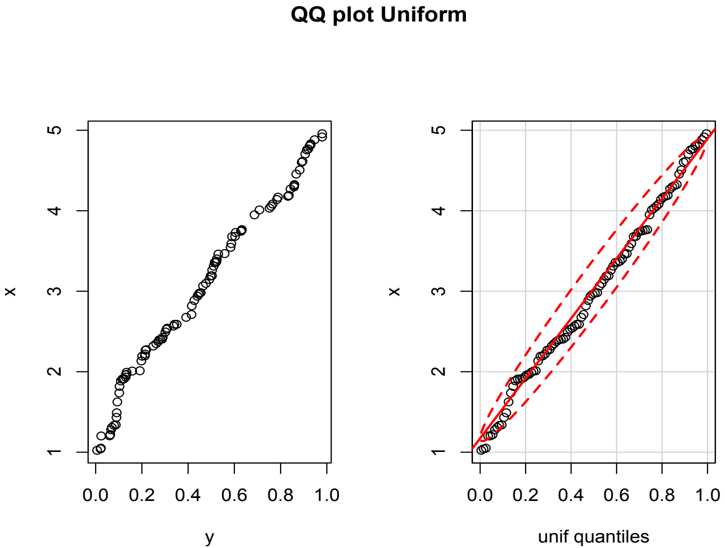


Graphical Representations

- Graphical uses of quantiles can be useful in determining aspects of the population from our random sample
- QQ plots: gives an indication of how close the ditribution of your random sample is to a theoretical distribution
  - called a normal QQ or normal probability plot when you compare to normal quantiles
  - QQ plot is similar to the EDF



Graphical Representations



Order Statistics
(CB 5.4)

Order Statistics

- Sample (empirical) quantiles are determined through order statistics.
- The order statistic of a random sample is denoted $X_{(1)}, X_{(2)}, ...X_{(n)}$ and satisfies $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ where $X_{(1)} = \min_{1 \leq i \leq n} X_i$
- For any number q between 0 and 1, the qth quantile is the observation that approximately nq of the observations are less than this observation and $n(1-q)$ are greater
- If nq is an integer, then the qth quantile is any real number such that $X_{(nq)} \leq X \leq X_{(nq+1)}$
- if nq is not an integer, then the qth quantile is $X_{\lceil nq \rceil}$ where $\lceil nq \rceil$ is the ceiling (smallest integer greater or equal to nq)
- The percentile is often used and is defined as the $100q$th sample percentile

Order Statistics

Example con't: quantiles of the empirical distribution

Recall our (ordered) random sample x={1.02, 1.16, 1.16, 1.93, 2.20, 2.25, 2.56, 2.58, 2.68, 3.08, 3.11, 3.55, 3.57, 3.91, 4.58, 4.80, 4.89, 4.92, 4.94, 4.99}

The median, $q = 0.5$ is any number between $x_{(10)} = 3.08$ and $x_{(11)} = 3.11$
The 25%ile, $q = 0.25$ is any number between $x_{(5)} = 2.20$ and $x_{(6)} = 2.25$
The 75%ile, $q = 0.75$ is any number between $x_{(15)} = 4.58$ and $x_{(16)} = 4.80$
The 99%ile $q = 0.99$ is $x_{(19.8)}$ which is $x_{(20)} = 4.99$ since $\lceil nq \rceil = \lceil 19.8 \rceil = 20$

Note: the population median (3) is not equal to the sample median $q = 0.5$ which is the mean of $x_{(10)} = 3.08$ and $x_{(11)} = 3.11$, $x = 3.095$

Order Statistics

- Note what we can have a non-unique median when nq is an integer
- This is commonly dealt with by the following:
- When n is odd then the empirical median is: $x_{\lceil n/2 \rceil}$
- When n is even then the empirical median is: $\frac{x_{(n/2)} + x_{n/2+1}}{2}$

Order Statistics: Discrete Distributions

- For a random sample $X_1, ..., X_n$ from a **discrete** distribution with pmf $f_X(x_i) = p_i$ and the possible values of X are in ascending order $x_1 < x_2 < ... < x_i$ then

$$
\begin{aligned}
P_0 &= 0 \\
P_1 &= p_1 \\
P_2 &= p_1 + p_2 \\
&\vdots \\
P_i &= p_1 + p_2 + ... + p_i
\end{aligned}
$$

- The order *statistics* from the sample are $X_{(1)}, X_{(2)}, ...X_{(n)}$, so:

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k}$$

and

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}]$$

Order Statistics: Discrete Distributions

- To prove $P(X_{(j)} \leq x_i)$, fix i and define Y to be a random variable that is the count of the number of $X_1, ..., X_n$ that are less than or equal to x_i
- Thus, the event $\{X_{(j)} \leq x_i\}$ can be thought of as a success and $\{X_{(j)} > x_i\}$ can be thought of as a failure
- With these definitions of success and failures, Y is defined as the number of successes in n trials. In other words, Y $\sim$ Bin(n,P$_i$)
- Relating back to our X's, the event $\{X_{(j)} \leq x_i\}$ is equivalent to the event $\{Y \geq j\}$ and we express this with the Binomal probability
- $P(X_{(j)} \leq x_i) = P(Y \geq j)$ and following this, the equality $P(X_{(j)} = x_i) = P(X_{(j)} \leq x_i) - P(X_{(j)} \leq x_{i-1})$. Thus the two equations are established.

Order Statistics: Discrete Distributions

Example: Probability of a discrete order random variable

Suppose we roll a dice 15 times (independent rolls), $P(X_i = x) = 1/6$. What is the probability that the third largest roll is at least 5?

We have the ordered random variables $X_{(1)}, ..., X_{(15)}$ with the third largest being the 13th of the 15 rolls. Thus, we want to find $P(X_{(13)} \geq 5)$.
From the definition $P_i = p_1 + p_2 + ... + p_i$, We have $P_i = P(x < 5) = 4/6$

$$\begin{aligned} P(X_{(13)} < 5) &= \sum_{k=13}^{15} \binom{15}{k} (4/6)^k (1 - 4/6)^{15-k} \\ &= 105(2/3)^{13}(1/3)^2 + 15(2/3)^{14}(1/3) + (2/3)^{15} \\ &= 0.07936 \end{aligned}$$

Thus $P(X_{(13)} \geq 5) = 1 - P(X_{(13)} < 5) = 1 - 0.07936 = 0.92064$

Order Statistics: Continuous Distributions

- For a random sample with order statistics $X_{(1)}, X_{(2)}, ...X_{(n)}$ from a **continuous** distribution with cdf $F_X(x)$ and pdf $f_X(x)$.
  The pdf of $X_{(j)}$ is:

$$f(X_{(j)}(x)) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j}$$

- The proof of this lies in taking the derivative of the cdf of $X_{(j)}$ to obtain the pdf (see CB theorem 5.4.4)
- As in the discrete case, define Y to be a random variable that is the count of the number of $X_1, ..., X_n$ that are less than or equal to x
- Thus, the event $\{X_{(j)} \leq x\}$ can be thought of as a success
- With this definition of success, Y is defined as the number of successes in n trials. In other words, Y $\sim$ Bin(n,F_x(x))
- Although X is continuous, by this definition Y is a counting variable and is discrete

Order Statistics: Continuous Distributions

- We dissect the pdf of the order statistic $(X_{(j)})$ $f(X_{(j)}(x))$ into three terms of interest:
  - $[F_X(x)]^{j-1}$ representing the j-1 sample items below x_i
  - $[1 - F_X(x)]^{n-j}$ representing the n-j sample items above x_i
  - f_X (x) representing the sample item near x_i

Example: Uniform Order Statistic

Suppose we have $X_{(1)}, X_{(2)}, ...X_{(5)}$ from a Uniform distribution on [0,1], what is the pdf of the the second order statistic? For Unif[0,1]:

$$f_X(x) = \begin{cases} 1, 0 \leq x \leq 1 \\ 0, \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0, x < 0 \\ x, 0 \leq x \leq 1 \\ 1, x > 1 \end{cases}$$

Order Statistics: Continuous Distributions

Example: Uniform Order Statistic, con't

$$f_{X_{(2)}}(x_2) = \frac{5!}{(2-1)!(5-2)!} f_X(x_2)[F_X(x_2)]^{2-1}[1-F_X(x_2)]^{5-2}$$
$$= \begin{cases} 20x_2(1-x_2)^3, 0 \leq x_2 \leq 1 \\ 0, \text{otherwise} \end{cases}$$

We also note that the jth order statistic from a uniform [0,1] has a beta(j, n-j+1) distribution

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-j}$$
$$= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} x^{j-1}(1-x)^{(n-j+1)-1}$$

From which the expected value and variance for the uniform order statistics can be defined: $E(X_{(j)}) = \frac{j}{n+1}$ and $\text{Var}(X_{(j)}) = \frac{j(n-j+1)}{(n+1)^2(n+2)}$

Order Statistics: Continuous Distributions

Joint distribution of order statistics will be discussed in class. Also see CB Theorem 5.4.6, specifically:
Let $X_{(1)}, ..., X_{(n)}$ be the order statistics for sample $X_1, ..., X_n$ from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of two order statistics $X_{(i)}$ and $X_{(j)}$ is

$$f_{X_{(i)},X_{(j)}}(u, v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) F_X(u)^{i-1} \times [F_X(v) - F_X(u)]^{j-1-i}[1 - F_X(v)]^{n-j}$$

The range is a special example of two order statistics.

Order Statistics: Continuous Distributions

Example: Exponential Distribution

Suppose a device has a lifetime X (hours) with pdf

$$f_X(x) = \begin{cases} 1/100 \exp^{-x/100}, x > 0 \\ 0, \text{otherwise} \end{cases}$$

1) Suppose there are two devices, and they operate independently but in series in a system. The system will fail when either device fails. Find the density function for the life of the system.
2) Suppose there are two devices, and they operate independently in parallel in a system. The system will fail when both fail. Find the density function for the life of the system.

Order Statistics: Continuous Distributions

Example: Uniform Distribution

Suppose a machine uses 10 batteries that have U[1/2, 1] distribution (in years), and it shuts off when 1/2 of the batteries are dead.
1) What is the expected time when the 5th battery will die?
2) What is the probability that the machine will shut off before 3/4 of a year?
3) The machine's efficiency is lost when there are 3 dead batteries. It costs $1 per day to run the machine at this point. How much money will the company spend before the machine shuts off? (Hint: This is a joint probability problem!)

Sampling Distributions Related to the Normal Distribution (CB 5.3)

Sampling Distributions

The derived distributions are **sampling distributions** related to the normal distribution. Recall, one of the goals of statistical inference is to estimate unknown population parameters by **statistics** (e.g. μ can be estimated by $\bar{X}$). The **probability distribution of a statistic** such as $\bar{X}$ is called the **sampling distribution**. These sampling distributions can be used to express uncertainty in the estimators. We find the sampling distributions of statistics using three possible methods:

1. Method of distribution functions (i.e. obtain CDF $F_X(x) = P(X \leq x)$, then differentiate $F_X(x)$ w.r.t. x to obtain the pdf $f_X(x)$).
2. Method of transformations. (CB Theorem 2.1.5)
3. Moment generating functions. (CB 2.3)

Sampling Distributions Related to the Normal Distribution

Under the assumption of normality, there are a few properties of $\bar{X}$ and S^2 that are important. First, recall for our sample,

$$\bar{X} = \frac{1}{n}\sum_{i-1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

So, for $X \sim (\mu, \sigma^2)$

1. $\bar{X}$ and S^2 are independent.
2. $\bar{X} \sim N(\mu, \sigma^2/n)$, namely $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/n$.
3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Sampling Distributions Related to the Normal Distribution

Proving 1, the independence between $\bar{X}$ and S^2, we look at n-1 deviations $(X_1 - \bar{X}, X_2 - \bar{X}, ..., X_{n-1} - \bar{X})$ and show that $\bar{X}$ is independent of $X_i - \bar{X}$ by showing $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0$. Since S^2 is a function of $X_i - \bar{X}$ then it is independent of $\bar{X}$.
The details of the proof are left as an exercise for Assignment 1.

Sampling Distribution of the Sample Mean of Normals

The proof of 2, $\bar{X} \sim N(\mu, \sigma^2/n)$ is straightforward. Because $X_1, X_2, ..., X_n$ is a random sample from a normal distribution with mean μ and variance σ^2, the X_i are independent, normally distributed variables with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$.

$$\begin{aligned}\bar{X} &= \frac{1}{n}\sum X_i = \frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n \\ &= a_1X_1 + a_2X_2 + ... + a_nX_n\end{aligned}$$

So $\bar{X}$ is a linear combination of $X_1, ..., X_n$. Recall that the sum of independently normally distributed random variables also has a normal distribution. Also, a linear transformation of a normally distributed variable is also normally distributed. This can be applied to conclude that $\bar{X}$ is normally distributed with:

$$E(\bar{X}) = E[\frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n] = \frac{1}{n}\mu + ... + \frac{1}{n}\mu = \mu$$

$$\text{Var}(\bar{X}) = \text{Var}[\frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n] = \frac{1}{n^2}\sigma^2 + ... + \frac{1}{n^2}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

Standard Normal Distribution and Z-statistic

Since $\bar{X}$ is normally distributed with mean μ and variance σ^2/n (sometimes denoted $\mu_{\bar{X}}$ and $\sigma^2_{\bar{X}}$) it follows that

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}(\frac{\bar{X} - \mu}{\sigma})$$

where Z has standard normal distribution.

Standard Normal Distribution and Z Random Variable

Example: The mean systolic blood pressure of adults with hypertension is μ mmHg (unknown population mean), with $\sigma^2 = 25$. A sample of $n = 10$ hypertensive patients at USC medical center is randomly selected, and the systolic blood pressure of each is measured. Find the probability that the sample mean will be within 1 mmHg of the true mean μ.

Standard Normal Distribution and Z Random Variable

Recall that if we increase the sample size, we approach the normal distribution via the central limit theorem. How many observations should be included in our sample if we wish $\bar{X}$ to be within 1 mmHg of the population mean with probability 0.95?

Z and χ^2

When $X_1, X_2, ..., X_n$ have the properties above, then $Z_i = (X_i - \mu)/\sigma$ are independent, standard normal random variables and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} (\frac{X_i - \mu}{\sigma})^2$$

has a χ^2 distribution with n degrees of freedom. The proof lies in that $X_1, X_2, ..., X_n$ is a random sample from a normal distribution with mean μ and variance σ^2, and the random variables Z_i are independent because the random variables X_i are independent. Moment generating functions are used to obtain the distribution of $\sum_{i=1}^{n} Z_i^2$.

χ^2 Distribution

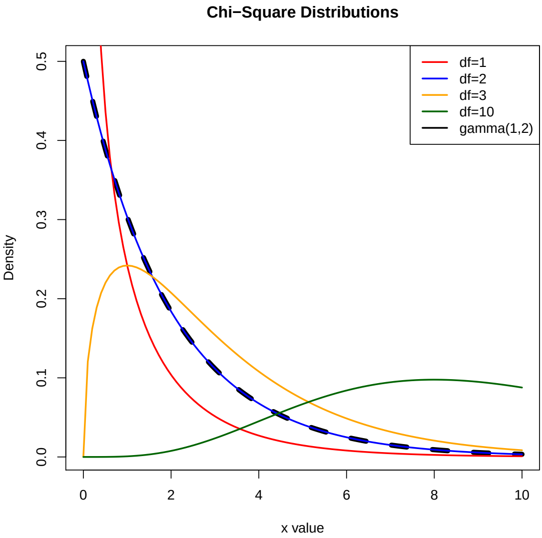
The general form of the χ^2 distribution with p degrees of freedom is

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}$$

Recall that the χ^2 distribution with p degrees of freedom is the same as that of the Gamma distribution with $\alpha = p/2$ and $\beta = 2$.

$$\chi^2_p \sim \text{Gamma}(\alpha = p/2, \beta = 2)$$

χ^2 Distribution



Sampling Distribution of the Sample Variance

For 3, we wish to determine the distribution of a multiple of the sample variance,

$$\bar{X} = \frac{1}{n}\sum_{i-1}^{n} X_i$$

$$S^2 = \frac{\sum_{i-1}^{n}(X_i - \bar{X})}{n-1}$$

then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

That is, $\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$.

Application of the Sampling Distribution of the Sample Variance

Example: Using the systolic blood pressure example above, we have a sample of n=10 and know $\sigma^2 = 25$. The 10 observations are used to calculate S^2, but suppose we want to find the two numbers such that the probability of S^2 being between those two numbers is 0.90, $P(b_1 \leq S^2 \leq b_2) = 0.90$

The t-distribution

The t-distribution (Student's t) is used when the population variance σ^2 is unknown. We use the estimate S^2, and the quantity

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

allows us to develop inferences about μ. This quantity has t-distribution with n-1 degrees of freedom.

Definition of t-distribution

If $U \sim N(0,1)$ and $V \sim \chi^2_p$, and U and V are independent, then

$$T = \frac{U}{\sqrt{V/p}}$$

Where T is a T statistic with t-distribution and p degrees of freedom.

The t-distribution

The pdf of T is:

$$f_T(t) = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})(p\pi)^{1/2}} \frac{1}{(1+\frac{t^2}{p})^{(p+1)/2}}$$

Proving this we start with the joint distribution of U and V

$$f_{U,V}(u,v) = \frac{1}{(2\pi)^{1/2}} e^{-u^2/2} \frac{1}{\Gamma(\frac{p}{2})2^{p/2}} v^{(p/2)-1} e^{-v/2}$$

and we apply the transformation

$$t = \frac{u}{\sqrt{v/p}},\ w = v$$

The t-distribution

This transformation maps f(u,v) to f(t,w). We apply the Jacobian transformation to do this. The Jacobian matrix is

$$\begin{vmatrix} \frac{du}{dt} & \frac{du}{dw} \\ \frac{dv}{dt} & \frac{dv}{dw} \end{vmatrix} = \begin{vmatrix} \sqrt{\frac{w}{p}} & \frac{t}{2}\sqrt{\frac{1}{wp}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{w}{p}}$$

We apply this transformation:

$$f_{T,W}(t,w) = f_{U,V}(u,v)|J|$$

Giving us

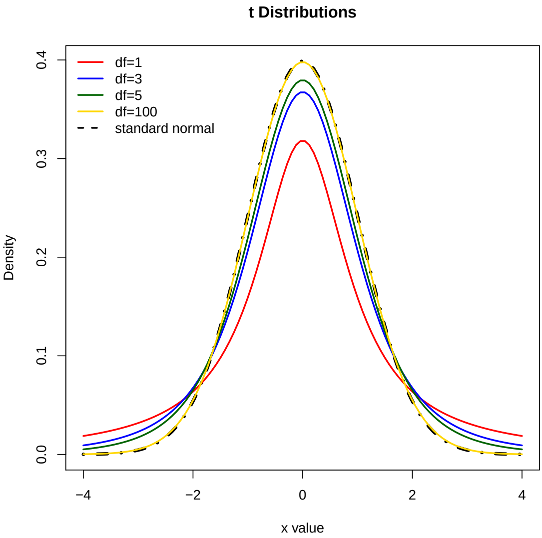
$$f_{T,W}(t,w) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{t^2 w}{2p}} \frac{1}{\Gamma(\frac{p}{2}) 2^{p/2}} w^{(p/2)-1} e^{-w/2} \sqrt{\frac{w}{p}}$$

The t-distribution

From this, the marginal density of T is obtained by

$$f_T(t) = \frac{1}{(2\pi)^{1/2}} \frac{1}{\Gamma(\frac{p}{2})2^{p/2}} \int_0^\infty e^{-\frac{t^2 w}{2p}} w^{(p/2)-1} e^{-w/2} \sqrt{\frac{w}{p}} dw$$
$$= \frac{1}{(2\pi)^{1/2}} \frac{1}{\Gamma(\frac{p}{2})2^{p/2} p^{1/2}} \Gamma(\frac{p+1}{2}) [\frac{2}{1+t^2/p}]^{(p+1)/2}$$

The t-Distribution



The F-distribution

The F-distribution is used when comparing variances of two normal populations based on random samples from those populations. We will have two samples of sizes n_1 and n_2 with variances σ_1 and σ_2 from which we calculate S_1 and S_2. The ratio S_1/S_2 is used to make inferences about the relative magnitudes of σ_1^2 and σ_2^2. If we divide each S_i^2 by σ_i^2 then the resulting ratio

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2}{\sigma_1^2}\frac{S_1^2}{S_2^2}$$

has F-distribution with $n_1 - 1$ numerator and $n_2 - 1$ denominator df.

Definition of F-distribution

If $U \sim \chi_p^2$ and $V \sim \chi_q^2$ are independent χ^2 random variables with p and q degrees of freedom, then

$$F = \frac{U/p}{V/q}$$

Where F is a F statistic with F-distribution and p numerator and q denominator degrees of freedom.

The F-distribution

To derive the F distribution we take the same approach as we did to find the t distribution. Start with the joint distribution of U and V (two χ^2 random variables), and make the Jacobian transformation. The joint distribution of U and V is:

$$
\begin{aligned}
f_U(u) &= \frac{1}{\Gamma(\frac{p}{2}2^{p/2})} u^{(p/2)-1} e^{-u/2} \\
f_V(v) &= \frac{1}{\Gamma(\frac{q}{2}2^{q/2})} v^{(q/2)-1} e^{-v/2} \\
f_{U,V}(u,v) &= \frac{1}{\Gamma(p)\Gamma(q)2^{p+q}} u^{p-1} v^{q-1} e^{-u/2} e^{-v/2}
\end{aligned}
$$

The F-distribution

The transformation is $f = u/v, , w = v$

$$\begin{vmatrix} \frac{du}{df} & \frac{du}{dw} \\ \frac{dv}{df} & \frac{dv}{dw} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ w & f \end{vmatrix} = -w$$

Applying the transformation and solving for the marginal distribution gives

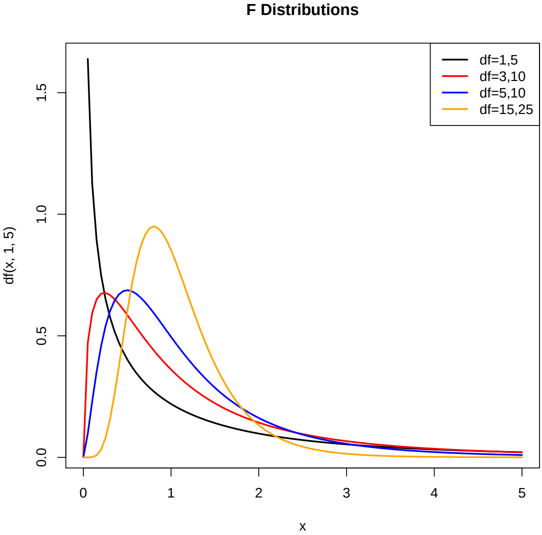
$$f_F(f) = \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})}(\frac{p}{q})^{p/2}\frac{x^{(p/2)-1}}{[1+(p/q)x]^{(p+q)/2}}$$

The F-distribution

Additional properties of the F-distribution:

1. If $X \sim F_{p,q}$, then $1/X \sim F_{q,p}$ (reciprocal also F random variable)
2. if $X \sim t_p$ then $X^2 \sim F_{1,p}$
3. if $X \sim F_{p,q}$ then $(p/q)X/(1+(p/q)X) \sim \text{beta}(p/2, q/2)$

The F-Distribution



Application of the F-distribution

Example: If we take two independent samples of size $n_1 = 6$ and $n_2 = 10$ from two normal populations with equal population variance, what value of the ratio of the two variances such that

$$P(\frac{S_1^2}{S_2^2} \leq b) = 0.95$$