Collections of Random Variables

Brian Vegetabile, Dustin Pluta

2018 Statistics Bootcamp
Department of Statistics
University of California, Irvine

September 14th, 2018

Considering Multiple Random Variables I

- While random variables are interesting in themselves, most of statistics revolves around collections of random variables.
- Need the tools to discuss the relationships between the random variables in a sample

Considering Multiple Random Variables II

- Ultimately we will want to make use of theorems and properties of random samples to perform inference on parameters of the distributions underlying a sample.
- For example, a statistic such as the sample mean will be used to make inference about the population mean
- We often rely on approximations that are due to "large" samples

Defining a Random Sample I

- We begin by defining a random sample

Definition (Random Sample)

Consider n random variables $X_1, X_2, \ldots, X_n$, these random variables form a **random sample** if each X_i is independent of all others and the marginal pmf or pdf of each RV is the same function f. Such random variables are said to be *independent and identically distributed*.

Definition (Sample Size)

The number of random variables, n, in a random sample is referred to as the sample size.

Defining a Random Sample II

- One way to think about the idea of the random sample is that there is some large or infinite ‘population’ where each random variable is selected from
- In this population, each random variable is generated using the same density or mass function f
- The sample size n is the number of random variables selected from that population.

Examples of Random Samples

- Consider 100 coin flips from a fair coin. Each coin flip can be considered as a random variable from a Bernoulli distribution with success probability $p = 0.5$.
- Consider the height of students in high schools around the country. It may be reasonable to assume that the heights of these students come from a population where height is represented as a normal distribution centered at some average μ with variance σ^2.
- Consider measuring 250 failure times for light bulbs from a single production line. The time to failure may be assumed to come from a population of light bulbs where failure time can be modeled as an exponential distribution with rate parameter λ.

The main idea is 'repeated observations' of the same phenomenon.

Joint Distribution of a Sample

- Based upon the definition of a random sample, we can construct the joint distribution which represents the probability distribution of the sample.
- Assuming a parameterized probability function $f(x_i|\theta)$ and we let $\mathbf{x} = (x_1, \ldots, x_n)$, then

$$f(\mathbf{x}|\theta) = f(x_1|\theta)f(x_2|\theta)\ldots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- Unfortunately this joint distribution will not always be nice to work with
- Additionally we may actually be interested in the distribution of a function of the random variables.

Defining Statistics & Sampling Distributions I

- Similar to the definition to random variable, we provide a few definitions

Definition (Statistic (DeGroot))

Suppose that the observable random variables of interest are $X_1, \ldots, X_n$. Let r be an arbitrary real-valued function of n real variables. Then the random variable $T = r(X_1, \ldots, X_n)$ is called a statistic.

Definition (Statistic (Dudewicz))

Any function of the random variables that are being observed say $t_n(X_1, X_2, \ldots, X_n)$, is called a statistic. Further since $X_1, X_2, \ldots, X_n$ are random variables, it is a random variable.

Defining Statistics & Sampling Distributions II

Definition (Statistic (Casella))

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, X_2, \ldots, X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a statistic. The probability distribution of a statistic Y is called the sampling distribution of Y.

Defining Statistics & Sampling Distributions III

- The major take away from all of these definitions is that a statistic is a function of random variables
- **Since it is a function of random variables, it is also a random variable and therefore has a distribution!**
- Often we will specifically be interested in the distribution of the statistic
- Since these distributions will often be unknown, we will look to finding approximations for them

Sums of Random Variables & the Sample Mean

- A primary statistic involves the sum of the random variables and functions of these sums
- That is

$$T(X_1, X_2, \ldots, X_n) = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

- The sample mean mean is a function of a sum of random variables,

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

Sample Variance

- Another statistic is the sample variance defined as

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

- The sample standard deviation is defined as $S = \sqrt{S^2}$.

Expectations of Sums of Random Variables I

- Claim: Let $X_1, X_2, \ldots, X_n$ be a random sample and let $g(x)$ be a function such that $E(g(X_1))$ and $Var(g(X_1))$ exist, then

$$E\left(\sum_{i=1}^{n} g(Xi)\right) = nE(g(X_1))$$

and

$$Var\left(\sum_{i=1}^{n} g(X_i)\right) = n(Var(g(X_1))$$

Expectations of Sums of Random Variables II

- Claim: Let $X_1, X_2, \ldots, X_n$ be a random sample and let $g(x)$ be a function such that $E(g(X_1))$ exists, then$E(\sum_{i=1}^{n} g(Xi)) = nE(g(X_1))$
- Demonstrate full proof
- Short "Proof":

$$E\left(\sum_{i=1}^{n} g(Xi)\right) = \sum_{i=1}^{n} E\left(g(Xi)\right) = nE(g(X_1))$$

Expectations of Sums of Random Variables III

- Now consider the sample mean, assume that the expectation of the population is μ and the variance in the population is σ^2
- This implies that

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}nE(X_1) \\ &= E(X_1) = \mu \end{aligned}$$

Expectations of Sums of Random Variables IV

- Additionally we can find the variance of the sample mean

$$\begin{aligned} Var(\bar{X}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2} n Var(X_1) \\ &= \frac{1}{n} Var(X_1) = \frac{\sigma^2}{n} \end{aligned}$$

- These two results imply that **regardless of the distribution of the statistic itself**, we know specific properties of the distribution!

MGF's for Sample Means I

- We also can calculate the MGF for the sample mean, specifically, define $Y = \frac{1}{n}\sum_{i=1}^{\infty} X_i$, where X_i are from a random sample (iid)...

$$\begin{aligned} M_Y(t) &= E(e^{tY}) \\ &= E\left(e^{\frac{t}{n}\sum_{i=1}^{n} X_i}\right) \end{aligned}$$

MGF's for Sample Means II

$$\begin{aligned}
M_Y(t) &= E(e^{\frac{t}{n}\sum_{i=1}^n X_i}) \\
&= \int \cdots \int\int e^{\frac{t}{n}\sum_{i=1}^n x_i} f(x_1, \ldots, x_n) dx_1 \ldots dx_n \\
&= \int \cdots \int\int \prod_{i=1}^n e^{\frac{t}{n} x_i} \prod_{i=1}^n f(x_i) dx_1 \ldots dx_n \\
&= [M_X(t/n)]^n
\end{aligned}$$

Example MGFs of the Sample Mean of Exponential Random Variables I

- The last result is mainly useful if we already know the distribution of the underlying observations of the sample.
- Consider a random sample of size n, $X_1, X_2, \ldots$ from a exponential distribution with MGF

$$M_X(t) = \frac{1}{1 - \lambda t}$$

Example MGFs of the Sample Mean of Exponential Random Variables II

therefore

$$M_X(t/n) = \left(\frac{1}{1 - \frac{\lambda}{n}t}\right)$$

and

$$M_{\bar{X}}(t) = \left(\frac{1}{1 - \frac{\lambda}{n}t}\right)^n$$

- Looking this up we see this is the MGF of a Gamma$(n, \frac{\lambda}{n})$
- We'll compare this with some approximations later

Bounding Probabilities on Statistics I

- The true distribution of a statistic is often difficult to obtain. Often, there are simple methods to get estimates for probability statements regarding the statistic
- Consider Markov's Inequality

Theorem (Markov's Inequality)

Suppose that X is a random variable such that $P(X \geq 0) = 1$, then for every real number $t > 0$

$$Pr(X \geq t) \leq \frac{E(X)}{t}$$

- DeGroot and Schervish state that the Markov inequality is primarily of interest for large values of t

Bounding Probabilities on Statistics II

- Related to Markov's Inequality and often more useful is Chebyshev's Inequality.

Theorem (Chebyshev's Inequality)

Let X be a random variable for which $Var(X)$ exists. Then for every real number $t > 0$,

$$Pr(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

- The proof follows from Markov's Inequality considering the random variable $Y = (X - E(X))^2$.

Bounding Probabilities on Statistics III

- If we consider the sample mean and apply Chebyshev's Inequality, it follows that

$$\Pr(|\bar{X} - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}$$

- This can be a very useful inequality for bounding probabilities, or helping to choose sample sizes

Application: Utilizing Chebyshev's Inequality

- From DeGroot and Schervish, Consider a random variable X with $Var(\sigma^2)$ and consider $t = 3\sigma$, then by Chebyshev's

$$\Pr(|X - E(X)| \geq 3\sigma) \leq \frac{\sigma^2}{(3\sigma)^2} = \frac{1}{9} \approx 0.11$$

- This implies that the probability that the a random variable will differ from its mean by more than 3 standard deviations is less than 0.11

Application: Chebyshev's and the Sample Mean I

DeGroot and Schervish Example 6.2.1 - Determining the Required Number of Observations

- Suppose that a random sample is to be taken from a distribution for which the value of the mean μ is not known, but for which it is known that the standard deviation σ is 2 units or less. We shall determine how large the sample size must be in order to make the probability at least 0.99 that $|\bar{X} - \mu|$ will be less than 1 unit.

Application: Chebyshev's and the Sample Mean II

- Since $\sigma^2 \leq 2^2 = 4$, it follows that for every sample of size n, that

$$\Pr(|\bar{X} - \mu| \geq 1) \leq \frac{4}{n}$$

Further by our problem statement we would like $\Pr(|\bar{X} - \mu| < 1) = 0.99$, thus $0.01 \leq \frac{4}{n}$ which implies that we need 400 observations.

General Idea of Asymptotics

- While the true distribution may be unknown for any given statistic, there may be assumptions for approximating the distribution 'as n grows large'
- That is we may be able to make some statements about the distribution of the statistic in the limit.
- This is where the Law of Large Numbers, Central Limit Theorem, and the Delta Method come into play
- We first review a few types of convergence.

Types of Convergence

- We'll be interested in the idea of what happens to the distribution of the statistic as the sample size grows to infinity.
- There are three main types of convergence we'll outline today
  - Convergence in Law (or Distribution)
  - Convergence in Probability
  - Almost-Sure Convergence

Convergence in Distribution

- We define our first mode of convergence: convergence in distribution or convergence in law

Definition (Convergence in Distribution)

A sequence of random variables, $X_1, X_2, \dots,$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for all points x where $F_X(x)$ is continuous. Denoted $X_n \xrightarrow{\mathcal{L}} X$ or $X_n \xrightarrow{\mathcal{D}} X$

- We see here that we are first talking about distribution functions converging to another distribution
- This is fundamentally different than the next few types of convergence.

Convergence in Probability and Almost Surely I

- A somewhat 'weak' convergence is outlined below

Definition (Convergence in Probability)

A sequence of random variables $X_1, X_2, \dots,$ converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

denoted $X_n \xrightarrow{P} X$.

- Notice that the form of this looks very similar to some of the inequalities that we have demonstrated...

Convergence in Probability and Almost Surely II

- A stronger version of convergence is almost-sure convergence

Definition (Almost-Sure Convergence)

A sequence of random variables $X_1, X_2, \ldots,$, converges almost surely to a random variable X if for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$$

denoted $X_n \xrightarrow{a.s.} X$.

- This is sometimes referred to as convergence with probability 1.

Example - Converges Almost Surely

Dudewicz and Mishra Example 6.2.6

- Let X_n be a sequence of random variables defined by

$$X_n = \begin{cases} 0 & \text{with probability } 1 - (\frac{1}{2})^n \\ 1 & \text{with probability } (\frac{1}{2})^n \end{cases}$$

for $n = 1, 2, 3, \ldots$. Then it can be shown that $P(\lim_{n\to\infty} X_n = 0) = 1$, hence $X_n \xrightarrow{a.s.} 0$.

Example - Converges in Probability I

Dudewicz and Mishra Example 6.2.6

- Let X_n be a sequence of random variables defined by

$$X_n = \begin{cases} 0 & \text{with probability } 1 - (\frac{1}{2})^n \\ 1 & \text{with probability } (\frac{1}{2})^n \end{cases}$$

for $n = 1, 2, 3, \ldots$. To show convergence in probability, we can appeal to Markov's Inequality...

Example - Converges in Probability II

- We see that $E(X_n) = (\frac{1}{2})^n$ and $E(X^2) = (\frac{1}{2})^n$.
- Therefore $Var(X_n) = \frac{2^n - 1}{2^{2n}}$
- Applying Markov's Inequality we have for every $\epsilon > 0$

$$\text{Pr}(|X_n| > \epsilon) \leq \frac{1}{2^n \epsilon^2}$$

and therefore

$$\lim_{n\to\infty} P(|X_n| \geq \epsilon) = 0$$

- Therefore the sequence X_n converges in probability to a random variable X that is "degenerate at zero" (takes on value 0 with probability 1)

Weak Law of Large Numbers I

- Understanding convergence principles will allow us to understand properties of the sampling distribution as the sample size grows.
- The first result of major importance is the Weak Law of Large Numbers

Theorem (Weak Law of Large Numbers)

Suppose that $X_1, \ldots, X_n$ form a random sample from a distribution for which the mean is μ and the variance exists. Let $\bar{X}_n$ denote the sample mean. Then

$$\bar{X}_n \xrightarrow{P} \mu$$

Weak Law of Large Numbers II

- Proof:

$$\text{Pr}(|\bar{X}_n - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2}$$

  Hence

$$\lim_{n\to\infty} \text{Pr}(|\bar{X}_n - \mu| < \epsilon) = 1$$

  showing the result.
- This result says that with high probability $\bar{X}_n$ tends to a value μ if the sample size is large
- This also suggests that if a large sample is taken of an unknown distribution, the sample mean will be a good approximation of the population mean with high probability.

Weak Law of Large Numbers R Demo

Central Limit Theorem I

- The WLLN is great start for understanding the distribution of the sample mean but fortunately, we can actually do better!

Theorem (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with mean μ and finite variance σ^2. Then

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} N(0, \sigma^2)$$

Central Limit Theorem II

- Sketch of proof: Use moment generating functions of characteristic functions to find the MGF of $Y = \sqrt{n}(\bar{X}_n - \mu)$. Expand this characteristic function using a Taylor series expansion and show that it converges to the MGF of a normal distribution with variance σ^2.
- There are additional versions of the Central Limit Theorem which reduce some of the assumptions, they will be introduced throughout the year.

Example - Distribution of the Sample Mean of Exponential Random Variables I

- The Central Limit Theorem may be one of the most important results in all of statistics.
- Consider the random sample $X_1, X_2, \ldots X_n$ where $X_i \sim Exponential(\lambda)$. That is

$$f(x|\lambda) = \frac{1}{\lambda} \exp\left\{-\frac{x}{\lambda}\right\}$$

Find the distribution of the sample mean from such a sample.

Example - Distribution of the Sample Mean of Exponential Random Variables II

- Appealing to the Central Limit Theorem, we know $E(X_i) = \lambda$ exists and further that the variance $Var(X) = \lambda^2$ is finite.
- This implies that

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{\mathcal{L}} N(0, \lambda^2)$$

- Thus we can say that

$$\bar{X}_n \dot{\sim} N\left(\lambda, \frac{\lambda^2}{n}\right)$$

Central Limit Theorem Example

Normal Approximation to the Binomial Distribution I

- In defining the binomial distribution, we stated that it could be thought of as the sum of independent Bernoulli trials with success probability p.
- We can attempt to approximate the Binomial distribution then using the Central Limit Theorem...
- First, $E(X) = p$ and $Var(X) = p(1-p)$, therefore

$$\sqrt{n}(\bar{X}_n - p) \xrightarrow{L} N(0, p(1-p))$$

based on the CLT

Normal Approximation to the Binomial Distribution II

- This implies that

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right) \xrightarrow{L} N(0, p(1-p))$$

factoring out a $1/n$, this becomes

$$\frac{\sqrt{n}}{n}\left(\sum_{i=1}^{n} X_i - np\right) \xrightarrow{L} N(0, p(1-p))$$

Normal Approximation to the Binomial Distribution III

- Now defining $Y = \sum_{i=1}^{n} X_i$ (A binomial random variable), we have

$$\frac{1}{\sqrt{n}}(Y - np) \xrightarrow{L} N(0, p(1-p))$$

- Rearranging, this implies that

$$Y \dot{\sim} N(np, np(1-p))$$

Normal Approximation to the Binomial Distribution IV

- Why is such an approximation useful?
- Recall the mass function for the binomial distribution

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

- $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ which can become computationally challenging for large n, where as the density of the normal distribution is relatively easy to calculate computationally.
- Therefore these approximations can become very useful throughout statistics

Normal Approximation to Binomial R Example

The Delta Method I

- The Central Limit Theorem is powerful in that it allows us to talk about the distribution of the sample mean.
- What if we're interested in more complicated functions of the sample mean?
- This is where the Delta Method comes into play
- We'll provide an informal derivation of the delta method

The Delta Method II

- Consider $X_1, X_2, \ldots, X_n$ which forms a random sample from a distribution that has a finite mean μ and finite variance σ^2.
- By CLT, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{L} N(0, \sigma^2)$.
- Now suppose there exists a function $g(\bar{X}_n)$ and we would like to approximate its distribution.

The Delta Method III

- The delta method works by taking a Taylor series expansion of $g(\bar{X}_n)$ at the mean of the distribution, that is

$$g(\bar{X}_n) \approx g(\mu) + g'(\mu)(\bar{X}_n - \mu) + \dots$$

and ignoring the higher order terms

The Delta Method IV

- Therefore

$$\begin{aligned} g(\bar{X}_n) - g(\mu) &= g'(\mu)(\bar{X}_n - \mu) \\ \sqrt{n}(g(\bar{X}_n) - g(\mu)) &= g'(\mu)\sqrt{n}(\bar{X}_n - \mu) \end{aligned}$$

- We know $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{L} N(0, \sigma^2)$, which implies that

$$\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{L} N(0, (g'(\mu))^2\sigma^2)$$

Example - Delta Method I

Ferguson - Chapter 7 Example 1

- Consider a random sample with mean μ and variance σ^2, by the $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{L} N(0, \sigma^2)$. What is the distribution of $\bar{X}_n^2$?

Example - Delta Method II

- Here $g(\bar{X}_n) = \bar{X}_n^2$, thus $g'(\bar{X}_n) = 2\bar{X}_n$, thus $g'(\mu) = 2\mu$. Utilizing the delta method formula we have that

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{L} N(0, 4\mu^2\sigma^2)$$

- Notice that if $\mu = 0$ this becomes a degenerate random variable and thus this approximation may not be useful...

References

Stephen Abbott. *Understanding analysis*. Springer, 2001.

George Casella and Roger L Berger. *Statistical inference*. Second edition, 2002.

EJ Dudewicz and SN Mishra. Modern mathematical statistics. john wilsey & sons. *Inc., West Sussex*, 1988.

Sheldon Ross. *A first course in probability*. Pearson, ninth edition, 2014.

Mark J Schervish. *Probability and Statistics*. 2014.