III. Random Samples and Modes of Convergence

MA481 – Mathematical Statistics

Academic Year 2014-2015

I'm afraid that I rather give myself away when I explain...Results without causes are much more impressive.

- Sherlock Holmes in The Stock-Broker's Clerk

1 Statistics

Example 1. *(Which Stat)* Suppose the time until an event occurs is modeled using an Exponential random variable with parameter λ. That is, $X_1, \ldots, X_n \overset{IID}{\sim} f_X$ where

$$f_X(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$$

If we are interested in estimating the mean λ, which would you prefer to use as an estimator: $\bar{X}$ or X_1? What about estimating the variance λ^2, would you prefer $\bar{X}^2$ or $S^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$?

- Let $X_1, \ldots, X_n$ be a random sample from a population, and let $T(\mathbf{x})$ be a real-valued (or vector-valued) function whose domain includes the support for $\mathbf{X}$. Then, the random variable (vector) $Y = T(\mathbf{X})$ is called a ________________.

- ⧫ Consider the case above, the distribution of Y is known as the ________________ ____________________ of Y.

Theorem 2. *(Mean and Variance of IID Sum) Let $X_1, \ldots, X_n$ be IID random variables and let g be a function such that $E\left[g\left(X_i\right)\right]$ exists. Then,*

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} g\left(X_i\right)\right) &= nE\left[g\left(X_1\right)\right] \\
Var\left(\sum_{i=1}^{n} g\left(X_i\right)\right) &= nVar\left[g\left(X_1\right)\right]
\end{aligned}
$$

Proof. □

Theorem 3. *(Sample Mean / Variance) Let $X_1, \ldots, X_n$ be IID random variables with mean μ and variance $\sigma^2 < \infty$. Then,*

- $E\left(\bar{X}\right) = \mu$
- $Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$
- $E\left(S^2\right) = \sigma^2$

where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and $S^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$.

Proof. □

- ♦ Let $X_1, \ldots, X_n$ be IID random variables with density $f_X(x)$. If the density is indexed by some parameter $\boldsymbol{\theta}$ we say that f_X is a ____________________ density and may highlight the dependence on the parameter by writing $f_X(x \mid \boldsymbol{\theta})$.
- ♦ Let $X_1, \ldots, X_n$ be IID random variables with parametric density $f_X(x \mid \boldsymbol{\theta})$, and let $T = T(\mathbf{X})$ be a statistic. We say that T is an ________________ __________________ for $\boldsymbol{\theta}$ if ________________________.

Example 4. *(Bias in S)* Let $X_1, \ldots, X_n$ be a random sample; show that S is a biased estimator for σ where S^2 is the usual unbiased estimator of σ^2.

Theorem 5. *(Jensen's Inequality) Let X be a random variable and g be a convex function, then*

$E[g(X)] \geq g[E(X)]$; if g is a concave function, then $E[g(X)] \leq g[E(X)]$. The inequality is strict if the function is strictly convex (or concave).

1.1 Order Statistics

Ordering the data is useful, as are the random variables that result from this process. They are especially helpful in studying extremes.

- The ________ ______________ of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted $X_{(1)}, \ldots, X_{(n)}$, where $X_{(i)} \leq X_{(j)}$ for all $i \leq j$.

Example 6. *(Density of Order Statistic)* Let $X_1, \ldots, X_n$ be IID random variables with CDF F and density f. Then, show that the density of $X_{(j)}$ is given by

$$f_{X_{(j)}}(x) = j\binom{n}{j} f(x)\left[F(x)\right]^{j-1}\left[1-F(x)\right]^{n-j}$$

where $j = 1, 2, \ldots, n$.

2 Modes of Convergence

Above, we were able to leverage the statistic and the underlying distribution to find particular sampling distributions. For example, we know the sampling distribution of an order statistic. We saw in previous notes and homework that the mean of a random sample from a Gamma distribution also has a Gamma distribution. But, these are special cases. What happens when the underlying population is unknown or does not result in a tractable solution?

- ► We can leverage the properties of IID random variables in order to obtain approximations to the _______________ ____________________ (or limiting distribution) when exact distributions are difficult to obtain analytically.

2.1 Convergence in Probability

Example 7. (Moving Uniform) Let $X_n \sim Unif(0, 1/n)$ and let $Y_n \sim Unif(0, 1 + 1/n)$. That is, the distribution of the random variable depends on the size of the sample from which it was drawn. Where are these random variables headed as n increases?

- ♦ A sequence of random variables $\{X_n\}$ is said to converge in ______________________ to a random variable X, written $X_n \xrightarrow{P} X$, if for every $\epsilon > 0$,

♠ ♠

- A useful result for establishing convergence in probability (though not always necessary) is Markov's Inequality.

Theorem 8. *(Markov's Inequality) Let X be a random variable and let g be a nonnegative function. Then, for any $\epsilon > 0$, we have*

$$Pr\left[g(X) \geq \epsilon\right] \leq \frac{E\left[g(X)\right]}{\epsilon}$$

Proof. □

Theorem 9. *(Weak Law of Large Numbers) Let $X_1, \ldots, X_n$ be IID random variables with mean μ. Define $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. Then, $\bar{X}_n \xrightarrow{P} \mu$.*

Proof. □

- Let $\mathbf{X}$ be a sample from a parametric distribution. Let $T_n(\mathbf{X})$ denote a statistic dependent on the sample size n. We say that T_n is a ____________________ estimator of the parameter θ if $T_n \xrightarrow{P} \theta$.

Theorem 10. *(Continuous Mapping Theorem) Suppose* $X_n \xrightarrow{P} X$. *If* g *is a continuous function, then* $g(X_n) \xrightarrow{P} g(X)$.

Example 11. *(Example 4 Cont.)* We have seen that the usual estimate of the standard deviation S is biased. However, show that S is a consistent estimator of σ.

2.2 Convergence in Distribution

Convergence in probability tells us where a statistic tends to be located, but it does not help us approximate probability statements about these statistics.

Example 12. *(Example 7 Cont.)* For the Moving Uniforms example, consider what happens to $Pr(X_n \leq x)$ and $Pr(Y_n \leq y)$ as n increases.

♦ A sequence of random variables $\{X_n\}$ is said to converge in ________________________ to a random variable X, written $X_n \stackrel{D}{\to} X$, if

♠ ♠

- Note that even though we say that X_n converges to X, the definition is actually a statement about the convergence of the distribution functions, not the random variables themselves.
- We often abuse notation and say something like $X_n \stackrel{D}{\to} N(0,1)$, meaning $X_n \stackrel{D}{\to} X$ where X follows a standard normal distribution.

Example 13. *(Minimum of Exponentials)* Let $X_i \sim Exp\,(\beta_i)$ for $i = 1, 2, \dots, n$. Suppose $\beta_i > 1$, what is the limiting distribution of $X_{(1)}$? Suppose $\beta_i = (1/2)^n$, what is the limiting distribution of $X_{(1)}$?

Theorem 14. *(Implications of Convergence) Consider a sequence of random variables $\{X_n\}$. If $X_n \stackrel{P}{\to} X$, then $X_n \stackrel{D}{\to} X$. The converse is only true if X is a random variable such that $Pr(X = a) = 1$ for some constant a.*

Theorem 15. *(Slutsky's Theorem) If $X_n \stackrel{D}{\to} X$, $Y_n \stackrel{P}{\to} a$, and $Z_n \stackrel{P}{\to} b$, then $Y_n X_n + Z_n \stackrel{D}{\to} aX + b$.*

Theorem 16. *(Central Limit Theorem) Let $X_1, \dots, X_n$ be IID random variables with mean μ and finite variance $\sigma^2 > 0$. Then*

$$\frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \stackrel{D}{\to} Z$$

where $Z \sim N(0, 1)$.

Example 17. *(One-Sample T-statistic)* Let $X_1, \ldots, X_n$ be IID random variables with mean μ and finite variance $\sigma^2 > 0$. Determine the limiting distribution of the one-sample t-statistic given by

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

where $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and $S^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$.