Math 502AB - Lecture 18

Dr. Jamshidian

October 25, 2017

1 Lecture - Part 1

1.1 Order Statistics, Cont'd

1.1.1 Quantiles

Definition

Let X be a continuous random variable with *cdf* $F(x)$. Then, for $0 < p < 1$, the p^{th} quantile for X is defined to be:

$$\xi_p = F^{-1}(p)$$

1.1.2 Sample Quantile

Strictly speaking, for $0 < p < 1$, the p^{th} sample quantile is the observation such that approximately np of the observations fall below it and $n(1-p)$ of the observations fall above it.

An example of **sample quantile** is, consider:

$$F_X\left(X_{(k)}\right) = P(X \leq X_{(k)}) \cong \text{ proportion of observations below } X_{(k)}$$

Essentially, if we know enough about this *cdf*, we can infer what the sample quantile might be. Consider:

$$\begin{aligned}
E\left[F\left(X_{(k)}\right)\right] &= \int_{-\infty}^{\infty} F(y)\frac{n!}{(k-1)!(n-k)!}[F(y)]^{k-1}[1-F(y)]^{n-k}f(y)dy \\
&\text{Now, let } Z = F(y) \Rightarrow dZ = f(y)dy \\
\Rightarrow &= \frac{n!}{(k-1)!(n-k)!}\int_0^1 Z^k(1-Z)^{n-k}dZ \\
\text{Since kernel of Beta dist'n} \quad &= \frac{n!\cdot k!}{(k-1)!(n-k)!}\cdot\frac{(n-k)!}{(n+1)!} = \frac{k}{n+1}
\end{aligned}$$

So, which order statistic do you expect to have the property that approximately a proportion p of the observations fall below it?

$$\frac{k}{n+1} = p \Rightarrow k = [p(n+1)]$$

And this is an *example* of a **sample quantile**.

1.2 Order Statistics for Discrete Random Variables

Let X be a *discrete* random variable, and let $X_1, ..., X_n$ be observations from X. We are interested in the distribution of $X_{(k)}$, or $P(X_{(k)} = x)$. The issue that we run into with *discrete distributions* is that there is a nonzero probability that two observations of X are equal to a, for some $a \in \mathcal{S}$.

We write:

$$P(X_{(k)} = x) = P(X_{(k)} \leq x) - P(X_{(k)} < x)$$

$$\begin{aligned} (1) \quad & P(X_{(k)} \leq x) = P(\text{no more than } n-k \text{ obs. fall above } x) \\ (2) \quad & P(X_{(k)} < x) = 1 - P(X_{(k)} \geq x) \\ & \qquad\qquad\quad = 1 - P(\text{No more than } k-1 \text{ obs. fall below x}) \end{aligned}$$

$$\begin{aligned} \text{We denote: } N_1 &= \# \text{ of observations} < x \\ N_2 &= \# \text{ of observations} = x \\ N_3 &= \# \text{ of observations} > x \end{aligned}$$

The distributions of these variables are *binomial*, with $P(X < x) = p_1$, $P(X = x) = p_2$, and $P(X > x) = p_3$. Now, we can show:

$$\begin{aligned} (1) \quad & P(X_{(k)} \leq x) = P(N_3 \leq n-k) \\ & \qquad\qquad\quad N_3 \sim Binomial(n, p_3) \\ (2) \quad & P(X_{(k)} < x) = 1 - P(N_1 \leq k-1) \end{aligned}$$

$$\Rightarrow P(X_{(k)} = x) = P(N_3 \leq n-k) + P(N_1 \leq k-1) - 1$$

Example:

1. Consider a random variable X with *pmf*:

$$f(x) = \frac{1}{5} \quad x = 1, 2, 3, 4, 5$$

Something like:

$$X = \begin{cases} 1 & \text{Freshman} \\ 2 & \text{Sophomore} \\ 3 & \text{Junior} \\ 4 & \text{Senior} \\ 5 & \text{Graduate} \end{cases}$$

We pick a random sample of size 20 students. What is the probability that $X_{(12)} = 3$?

$$\begin{aligned} P(X_{(12)} = 3) &= P(N_3 \leq 20 - 12) + P(N_1 \leq 12 - 1) - 1 \\ &= pbinom(20 - 12, 20, 0.4) + pbinom(12 - 1, 20, 0.4) - 1 \end{aligned}$$

1.3 Convergence Concepts in Probability

Consider a sequence of random variables:

$$\{X_n\}_{n=1}^{\infty}$$

We are interested in investigating its behavior as $n \to \infty$. To look at these properties, let's start with sequences of real numbers.

Consider a sequence:

$$\{x_n\}_{n=1}^{\infty} \quad x_n \in \mathbb{R}$$

If we had $x_n \to a$, what does this mean? Well, we go back to our analysis definition where $\forall \epsilon > 0$, $\exists N_0$ such that $\forall n \geq N_0$, we have $|x_n - a| < \epsilon$. But this method *will not* work with random variables. Because of this, we need different approaches.

1.3.1 Convergence in Probability

Definition:

We say that a sequence of random variables $\{X_n\}_{n=1}^{\infty}$ converges in probability to α if:

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

Or, equivalently:

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

1.3.2 Weak Law of Large Numbers (WLLN)

Let $X_1, ..., X_n$ be a sample from a random variable with mean μ and variance $\sigma^2 < \infty$. Let

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{1}$$

Then

$$\overline{X}_n \xrightarrow{P} \mu \quad \text{as } n \to \infty$$

Recall that $E[\overline{X}_n] = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$.

$$\begin{aligned} P\left(|\overline{X}_n - \mu| > \epsilon\right) &\le \frac{Var(\overline{X}_n)}{\epsilon^2} \text{ , by Chebychev} \\ &= \frac{\sigma^2}{n\epsilon^2} \end{aligned}$$

$$\Rightarrow \lim_{n\to\infty} P(|\overline{X}_n - \mu| > \epsilon) = 0$$

Example:

1. **Application -** *Monte-Carlo Integration*: Suppose that we wished to obtain the value of:

$$I = \int_0^1 g(x)dx$$

for some function $g(x)$. Let $X \sim Unif(0,1)$. Then:

$$E[g(x)] = \int_0^1 g(x)dx$$

On the other hand, by the **Weak Law of Large Numbers**:

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \xrightarrow{P} E[g(X_i)]$$

So, we can use this property to get our total (Computer example given in class).

2. **WLLN for** S_n^2

Let $X_1, ..., X_n$ (**iid**), with:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Then:

$$P(|S_n^2 - \sigma^2| \ge \epsilon) \le \frac{E\left[(S_n^2 - \sigma^2)^2\right]}{\epsilon^2} = \frac{Var(S_n^2)}{\epsilon^2}$$

So, $S_n^2 \xrightarrow{P} \sigma^2$ if $Var(S_n^2) \to 0$ as $n \to \infty$.

2 Lecture - Part 2

2.1 Convergence in Probability, Cont'd

2.1.1 Theorem:

Suppose that $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$

Proof:

$$|(X_n + Y_n) - (X + Y)| \leq |X_n - X| + |Y_n - Y|$$
$$\Rightarrow P\left[|X_n - X| + |Y_n - Y| \geq \epsilon\right] \geq P\left[|(X_n + Y_n) - (X + Y)| \geq \epsilon\right]$$

On the other hand, using the result that if $U > 0$, $V > 0$:

$$P(U + V \geq \epsilon) \leq P\left(U \geq \frac{\epsilon}{2}\right) + P\left(V \geq \frac{\epsilon}{2}\right)$$

We can then write:

$$\begin{aligned} P\left[|(X_n + Y_n) - (X + Y)| \geq \epsilon\right] &\leq P\left[|X_n - X| + |Y_n - Y| \geq \epsilon\right] \\ &\leq P\left(|X_n - X| \geq \frac{\epsilon}{2}\right) + P\left(|Y_n - Y| \geq \frac{\epsilon}{2}\right) \end{aligned}$$

Take limits as $n \to \infty$, and you are done.

2.1.2 Theorem:

Suppose that $X_n \to a$, and $h(x)$ is a continuous function at a. Then $h(X_n) \to h(a)$

Proof:

Since h is continuous at $x = a$, then for every $\epsilon > 0$, there $\exists \delta > 0$, such that $|x - a| < \delta \Rightarrow |h(x) - h(a)| < \epsilon$ (or, equivalently, $|h(x) - h(a)| > \epsilon \Rightarrow |x - a| < \delta$. Substituding X_n for x, we have:

$$|h(X_n) - h(a)| > \epsilon \Rightarrow |X_n - a| > \delta$$

Because of this condition (ie. the left side is a subset of the rightside), we have:

$$P(|h(X_n) - h(a)| > \epsilon) \leq P(|X_n - a| < \delta)$$

Since the right hand side converges to zero, the left must be zero as well and we are done.

2.1.3 Theorem

Suppose $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$. Then $X_n Y_n \xrightarrow{P} XY$.

Proof:

$$X_nY_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2 \xrightarrow{P} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X - Y)^2 = XY$$

2.1.4 Examples

1. Let $X_1, ..., X_n$ be a sample from a population with mean μ, variance σ^2, and $E[X_1^4] < \infty$. Then $S_n^2 \xrightarrow{P} \sigma^2$.

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\ &= \left(\frac{n}{n-1}\right)\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\overline{X}_n\right)^2\right) \\ &\xrightarrow[n\to\infty]{P} 1 \cdot \left(E[X_1^2] - (E[X_1])^2\right) = \sigma^2 \end{aligned}$$

2.2 Convergence in Distribution

Definition

Let $X_1, X_2, ...$ be a sequence of random variables with *cdf*s $F_{X_1}, F_{X_2}, ...,$ respectively. Let X be a random vairable with cdf F_X. We say that X_n **Converges in Distribution** to X if

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

at every point at which X is continuous.

Example

1. Consider:

$$F_{X_n}(x) = \begin{cases} 1 & x \geq \frac{1}{n} \\ 0 & \text{otherwise} \end{cases}$$

This means, we have:

$$F_{X_n}(x) \xrightarrow{n\to\infty} F_X(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Note that, though $F_{X_n}(0) = 0$, by definition:

$$\lim_{n\to\infty} F_{X_n}(0) = F_X(0) = 1$$

2. Let $X_1, ..., X_n$ be a sample from $Unif(0, 1)$. Define $Y_n = X_{(n)}$. We know that:

$$F_{Y_n}(y) = \begin{cases} 0 & \text{if } y < 0 \\ y^n & \text{if } 0 < y \leq 1 \\ 1 & \text{if } y > 1 \end{cases}$$

But, as $n \to \infty$, $Y_n \xrightarrow{D} 1$:

$$\begin{cases} 1 & y \geq 1 \\ 0 & y < 1 \end{cases}$$

This is referred to as a *degenerate distribution*, that is, a distribution where all mass is at one point.

3. Suppose that $X_1, ..., X_n$ is a sample from the *Pareto distribution*, with *cdf*:

$$F_{X_i}(x) = 1 - \frac{1}{1+x} \quad x > 0$$

Define $Y_n = nX_{(1)}$. It can be shown that:

$$\begin{aligned} F_{Y_n}(y) &= 1 - \left(1 + \frac{y}{n}\right)^{-n} \\ \lim_{n\to\infty} F_{Y_n}(y) &= 1 - e^{-y} \\ Y_n &\xrightarrow{D} Y \sim exp(1) \end{aligned}$$

4. Let $X_1, ..., X_n \sim Bernouli(p)$. We have:

$$Y_n = \sum_{i=1}^{n} X_i \sim Binomial(n, p)$$

Consider $W_n = \frac{Y_n}{n}$. Show that $W_n \xrightarrow{D} p$. To show this, consider the moment generating functions:

$$\begin{aligned} M_{W_n}(t) = M_{Y_n}\left(\frac{t}{n}\right) &= \left(pe^{t/n} + 1 - p\right)^n \quad \text{(the MGF for binomial distn.)} \\ &= \left[p\left(1 + \frac{t}{n} + \frac{t^2}{2n^2} + \cdots\right) + 1 - p\right]^n \\ &= \left[1 + \frac{pt}{n} + \mathcal{O}\left(\frac{1}{n^2}\right)\right]^n \to e^{pt} \end{aligned}$$

Since we know that e^{pt} is the *moment generating function* for a degenerate random variable of a single point.

2.2.1 The Central Limit Theorem

Let $X_1, ..., X_n$ be a set of **iid** random variables with mean μ and variance σ^2. Then,

$$\frac{\sum X_i - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{Dist.} N(0, 1)$$

Proof:

Let

$$Z_n = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i, \quad \text{where } Y_i = \frac{X_i - \mu}{\sigma}$$

We need to show that:

$$M_{Z_n}(t) = \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n \xrightarrow{n\to\infty} e^{t^2/2}$$

It is sufficient to show that:

$$\log M_{Z_n}(t) \xrightarrow{n\to\infty} \frac{t^2}{2}$$

Let $L(t) = \log\left(M_{Y_1}(t)\right)$. Then:

$$L\left(\frac{t}{\sqrt{n}}\right) = L(0) + L'(0)\frac{t}{\sqrt{n}} + L''(0)\frac{t^2}{2n} + \mathcal{O}\left(\frac{t}{\sqrt{n}}\right)^3$$

$$\begin{aligned}
L(0) &= \log\left(M_{Y_1}(0)\right) = 0 \\
L'(t) = \frac{M'_{Y_1}(t)}{M_{Y_1}(t)} \Rightarrow L'(0) &= \frac{M'_{Y_1}(0)}{M_{Y_1}(0)} = \frac{E[Y_1]}{1} = 0 \\
L''(t) &= \frac{M''_{Y_1}(t)M_{Y_1}(t) - \left[M'_{Y_1}(t)\right]^2}{\left(M_{Y_1}(t)\right)^2} \\
\Rightarrow L''(0) &= \frac{M''_{Y_1}(0)M_{Y_1}(0) - \left[M'_{Y_1}(0)\right]^2}{\left(M_{Y_1}(0)\right)^2} = \frac{M''_{Y_1}(0) = 1 - 0}{1} \\
&= M''_{Y_1}(0) = E\left[Y_1^2\right] = 1
\end{aligned}$$

Now we look at:

$$\begin{aligned}
n \log M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) &= nL\left(\frac{t}{\sqrt{n}}\right) \\
&= n\left[L(0) + L'(0)\frac{t}{\sqrt{2n}} + L''(0)\frac{t^2}{n} + \mathcal{O}\left(\frac{t}{\sqrt{n}}\right)^3\right] \\
&= n\left[0 + 0 + 1\cdot\frac{t^2}{2n} + \mathcal{O}\left(\frac{t}{\sqrt{n}}\right)^3\right] \\
&= \frac{t^2}{2} + \mathcal{O}\left(\frac{1}{n^{1/2}}\right) \to \frac{t^2}{2}
\end{aligned}$$