CHAPTER 5

5. PROBABILITY RESULTS & LIMIT THEOREMS

5.1 BOUNDS ON PROBABILITIES BASED ON MOMENTS

THEOREM

If X is a random variable, and h is a non-negative real function, then for any $c > 0$,

$$\mathrm{P}[\ h(X) \geq c\ ] \leq \frac{\mathrm{E}_{f_X}[\ h(X)\ ]}{c}$$

Proof (continuous case) : Suppose that X has density function f_X which is positive for $x \in \mathbb{X}$. Let $\mathcal{A} = \{\ x \in \mathbb{X} :\ h(x) \geq c\} \subseteq \mathbb{X}$. Then, as $h(x) \geq c$ on $\mathcal{A}$,

$$\begin{aligned}
\mathrm{E}_{f_X}[\ h(X)\ ] &= \int_{\mathbb{X}} h(x) f_X(x)\ dx = \int_{\mathcal{A}} h(x) f_X(x)\ dx + \int_{\mathcal{A}'} h(x) f_X(x)\ dx \\
&\geq \int_{\mathcal{A}} h(x) f_X(x)\ dx \geq \int_{\mathcal{A}} c f_X(x)\ dx = c\,\mathrm{P}[\ X \in \mathcal{A}\ ] = c\mathrm{P}[\ h(X) \geq c\ ]
\end{aligned}$$

Special Case : The **Markov Property** : $h(x) = |x|^r$ for $r > 0$, so $\mathrm{P}[\ |X|^r \geq c\ ] \leq \dfrac{\mathrm{E}_{f_X}[\ |X|^r\ ]}{c}$

THEOREM (**CHEBYCHEV INEQUALITY**)

Suppose that X is a random variable with expectation μ and variance σ^2. Then for any $k > 0$,

$$\mathrm{P}[\ |X - \mu| \geq k\sigma\ ] \leq \frac{1}{k^2}$$

Proof : Put $h(x) = (x - \mu)^2$ and $c = k^2\sigma^2$ in the previous theorem.

Corollary For $\epsilon > 0$, $\mathrm{P}[\ |X - \mu| \geq \epsilon\ ] \leq \dfrac{\sigma^2}{\epsilon^2}$ $\quad$ $\mathrm{P}[\ |X - \mu| < \epsilon\ ] \geq 1 - \dfrac{\sigma^2}{\epsilon^2}$

5.2 A BOUND ON EXPECTED VALUES

DEFINITION

A function g is **convex** if $\dfrac{d^2}{dt^2}\{g(t)\}_{t=x} = g''(x) > 0$ for all x.

THEOREM (**JENSEN'S INEQUALITY**)

Suppose that X is a random variable, and function g is convex. Then

$$\mathrm{E}_{f_X}[\ g(X)\ ] \geq g(\ \mathrm{E}_{f_X}[\ X\ ]\ )$$

Proof : A Taylor expansion of $g(x)$ around $x = \mu$ gives $g(x) = g(\mu) + (x - \mu)g'(\mu) + \frac{1}{2}(x - \mu)^2 g''(x_0)$ for some x_0 such that $x < x_0 < \mu$. Thus, taking expectations,

$$\mathrm{E}_{f_X}[\ g(X)\ ] = g(\mu) + \mathrm{E}_{f_X}[\ (X - \mu)\ ] g'(\mu) + \frac{1}{2}\mathrm{E}_{f_X}\left[\ (X - \mu)^2\ \right] g''(x_0) \geq g(\mu) = g(\ \mathrm{E}_{f_X}[\ X\ ]\ )$$

as $\mathrm{E}_{f_X}[\ (X - \mu)\ ] = 0$, and $\mathrm{E}_{f_X}\left[\ (X - \mu)^2\ \right],\ g''(x_0) \geq 0$.

5.3 CONVERGENCE IN DISTRIBUTION

DEFINITION

Consider a sequence of random variables $X_1, X_2, ...$ and a corresponding sequence of cdfs, $F_{X_1}, F_{X_2}, ...$ so that for $n = 1, 2, ..$ $F_{X_n}(x) = \mathrm{P}[\ X_n \leq x\ ]$. Suppose that there exists a cdf, F_X, such that **for all** x **at which** F_X is **continuous**,

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

Then $X_1, ..., X_n$ **converges in distribution** to random variable X with cdf F_X, denoted

$$X_n \xrightarrow{d} X$$

and F_X is the **limiting distribution**.

Note: Convergence of a sequence of mgfs also indicates convergence in distribution, that is, if for all t at which $M_X(t)$ is defined, if as $n \longrightarrow \infty$, we have

$$M_{X_i}(t) \longrightarrow M_X(t) \qquad \text{then} \qquad X_n \xrightarrow{d} X$$

- this result was used to prove that the Poisson distribution arose as the limiting case of the Binomial distribution, and also to prove the Central Limit Theorem.

DEFINITION

The sequence of random variables $X_1, ..., X_n$ converges in distribution to constant c if the limiting distribution of $X_1, ..., X_n$ is **degenerate at** c, that is,

$$X_n \xrightarrow{d} X$$

and $\mathrm{P}[\ X = c\ ] = 1$, so that

$$F_X(x) = \begin{cases} 0 & x < c \\ \\ 1 & x \geq c \end{cases}$$

Interpretation A *special case* of convergence in distribution occurs when the limiting distribution is *discrete*, with the probability mass function only being non-zero at a *single* value, that is, if the limiting random variable is X, then

$$f_X(x) = 1 \qquad x = c, \text{ and zero otherwise.}$$

The following theorem illustrates another aspect of convergence in distribution.

THEOREM

The sequence of random variables $X_1, ..., X_n$ converges in distribution to c if and only if, for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} \mathrm{P}[\ |X_n - c| < \epsilon\ ] = 1$$

This theorem indicates that convergence in distribution to a constant c occurs if and only if the probability becomes increasingly concentrated around c as $n \longrightarrow \infty$.

5.4 CONVERGENCE IN PROBABILITY

DEFINITION (CONVERGENCE TO A CONSTANT)

The sequence of random variables $X_1, ..., X_n$ **converges in probability** to **constant** c, denoted

$$X_n \xrightarrow{p} c$$

if

$$\lim_{n \longrightarrow \infty} \mathrm{P}[\ |X_n - c| < \epsilon\ ] = 1 \qquad \text{or equivalently} \qquad \lim_{n \longrightarrow \infty} \mathrm{P}[\ |X_n - c| \geq \epsilon\ ] = 0$$

that is, if the limiting distribution of $X_1, ..., X_n$ is degenerate at c.

Interpretation Convergence in probability to a constant is *precisely* equivalent to convergence in distribution to a constant.

THEOREM (**WEAK LAW OF LARGE NUMBERS**)

Suppose that $X_1, ..., X_n$ is a sequence of i.i.d. random variables with expectation μ and variance σ^2. Let Y_n be defined by

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

then, for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} \mathrm{P}[\ |Y_n - \mu| < \epsilon\ ] = 1,$$

that is, $Y_n \xrightarrow{p} \mu$, and thus the **mean** of $X_1, ..., X_n$ converges in probability to μ.

Proof

Using the properties of expectation, it can be shown that Y_n has expectation μ and variance σ^2/n, and hence by the **Chebychev Inequality**,

$$\mathrm{P}[\ |Y_n - \mu| \geq \epsilon\ ] \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0 \quad \text{as } n \longrightarrow \infty$$

for all $\epsilon > 0$. Hence

$$\mathrm{P}[\ |Y_n - \mu| < \epsilon\ ] \longrightarrow 1 \quad \text{as } n \longrightarrow \infty$$

DEFINITION (CONVERGENCE TO A RANDOM VARIABLE)

The sequence of random variables $X_1, ..., X_n$ **converges in probability** to **random variable** X, denoted $X_n \xrightarrow{p} X$, if, for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} \mathrm{P}[\ |X_n - X| < \epsilon\ ] = 1 \qquad \text{or equivalently} \qquad \lim_{n \longrightarrow \infty} \mathrm{P}[\ |X_n - X| \geq \epsilon\ ] = 0$$

THEOREM

For sequence of random variables $X_1, ..., X_n$ if

$$X_n \xrightarrow{p} X$$

then

$$X_n \xrightarrow{d} X$$

so convergence in probability to a random variable implies convergence in distribution.

5.5 ALMOST SURE CONVERGENCE

DEFINITION

The sequence of random variables $X_1, ..., X_n$ **converges almost surely** to **constant** c, denoted

$$X_n \xrightarrow{a.s.} c$$

if

$$\mathrm{P}\left[\lim_{n \longrightarrow \infty} |X_n - c| < \epsilon\right] = 1$$

Interpretation Recall the fundamental definition of a random variable as a real-valued function from sample space Ω to $\mathbb{R}$. The sequence of random variables $X_1, ...X_n$ corresponds to a sequence of functions defined on elements of Ω. Almost sure convergence requires that the sequence of real numbers $X_n(\omega)$ converges to c (as a real sequence) for all $\omega \in \Omega$, as $n \longrightarrow \infty$, except perhaps when ω is in a set having probability zero.

THEOREM (STRONG LAW OF LARGE NUMBERS)

Suppose that $X_1, ..., X_n$ is a sequence of i.i.d. random variables with expectation μ and (finite) variance σ^2. Let Y_n be defined by

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

then, for all $\epsilon > 0$,

$$\mathrm{P}\left[\lim_{n \longrightarrow \infty} |Y_n - \mu| < \epsilon\right] = 1,$$

that is, $Y_n \xrightarrow{a.s.} \mu$, and thus the **mean** of $X_1, ..., X_n$ converges almost surely to μ.

DEFINITION

The sequence of random variables $X_1, ..., X_n$ **converges almost surely** to **random variable** X, denoted $X_n \xrightarrow{a.s.} X$, if, for all $\epsilon > 0$,

$$\mathrm{P}\left[\lim_{n \longrightarrow \infty} |X_n - X| < \epsilon\right] = 1$$

Interpretation As above, recalling that random variable is a real-valued function from sample space Ω to $\mathbb{R}$, almost sure convergence to random variable X requires that the sequence of real numbers $X_n(\omega)$ converges to $X(\omega)$ (as a real sequence) for all $\omega \in \Omega$, as $n \longrightarrow \infty$, except perhaps when ω is in a set having probability zero.

THEOREM

For sequence of random variables $X_1, ..., X_n$

$$X_n \xrightarrow{a.s.} X \quad \Longrightarrow \quad X_n \xrightarrow{p} X \quad \Longrightarrow \quad X_n \xrightarrow{d} X$$

so almost sure convergence implies convergence in probability, which implies convergence in distribution to random variable X.