CHAPTER 5

PROBABILITY RESULTS & LIMIT THEOREMS

5.1 BOUNDS ON PROBABILITIES BASED ON MOMENTS

Theorem 5.1.1 *If X is a random variable, and h is a non-negative real function, then for any $c > 0$,*

$$P\left[h(X) \geq c\right] \leq \frac{E_{f_X}\left[h(X)\right]}{c}$$

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

$$\begin{aligned} E_{f_X}\left[h(X)\right] &= \int_{\mathbb{X}} h(x) f_X(x) dx = \int_{\mathcal{A}} h(x) f_X(x) dx + \int_{\mathcal{A}'} h(x) f_X(x) dx \\ &\geq \int_{\mathcal{A}} h(x) f_X(x) dx \geq \int_{\mathcal{A}} c f_X(x) dx = cP\left[X \in \mathcal{A}\right] = cP\left[h(X) \geq c\right] \end{aligned}$$

Special Case : The **Markov Property** : $h(x) = |x|^r$ for $r > 0$, so

$$P\left[|X|^r \geq c\right] \leq \frac{E_{f_X}\left[|X|^r\right]}{c}$$

Theorem 5.1.2 ***THE CHEBYCHEV INEQUALITY***

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

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

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

Corollary For $\epsilon > 0$,

$$P\left[|X - \mu| \geq \epsilon\right] \leq \frac{\sigma^2}{\epsilon^2} \qquad \text{and} \qquad P\left[|X - \mu| < \epsilon\right] \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

5.1.1 A BOUND ON EXPECTED VALUES*

Definition 5.1.1 A function g is convex if, for all x, $\dfrac{d^2}{dt^2}\{g(t)\}_{t=x} = g''(x) > 0$

Theorem 5.1.3 ***JENSEN'S INEQUALITY***

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

$$E_{f_X}[g(X)] \geq g(E_{f_X}[X])$$

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,

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

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

5.2 CONVERGENCE FOR PROBABILITY MODELS

5.2.1 THE CENTRAL LIMIT THEOREM

Theorem 5.2.1 *Suppose $X_1, ..., X_n$ are i.i.d. random variables with mgf M_X, with*

$$E_{f_X}[X_i] = \mu \qquad Var_{f_X}[X_i] = \sigma^2$$

both finite. Let the random variable Z_n be defined by

$$Z_n = \frac{\displaystyle\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

and let Z_n have mgf M_{Z_n}. Then, as $n \longrightarrow \infty$,

$$M_{Z_n}(t) \longrightarrow \exp\left\{\frac{t^2}{2}\right\}$$

irrespective *of the form of M_X.*

Proof. First, let $Y_i = (X_i - \mu)/\sigma$ for $i = 1, ..., n$. Then $Y_1, ..., Y_n$ are i.i.d. with mgf M_Y say, and by the elementary properties of expectation, $\mathrm{E}_{f_Y}[Y_i] = 0$,$\mathrm{Var}_{f_Y}[Y_i] = 1$ for each i. Using the power series expansion result for mgfs, we have that

$$M_Y(t) = 1 + tE_{f_Y}[Y] + \frac{t^2}{2!}E_{f_Y}[Y^2] + \frac{t^3}{3!}E_{f_Y}[Y^3] + ... = 1 + \frac{t^2}{2!} + E_{f_Y}[Y^3] + .$$

Now, the random variable Z_n can be rewritten

$$Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)$$

and thus, again by a standard mgf result, as $Y_1, ..., Y_n$ are independent, we have that

$$M_{Z_n}(t) = \prod_{i=1}^{n} \left\{ M_Y\left(\frac{t}{\sqrt{n}}\right) \right\} = \left\{ 1 + \frac{t^2}{2n} + \frac{t^3}{6n^{3/2}} E_{f_Y}[Y^3] + ... \right\}^n$$

Taking logs, and using the expansion $\log(1+s) = s - s^2/2 + s^3/3 - ...$ we have that

$$\log M_{Z_n}(t) = n \left[\left(\frac{t^2}{2n} + \frac{t^3}{6n^{3/2}} E_{f_Y}[Y^3] + ... \right) - \frac{1}{2} \left(\frac{t^2}{2n} + \frac{t^3}{6n^{3/2}} E_{f_Y}[Y^3] + ... \right)^2 + ... \right]$$

Thus, as $n \longrightarrow \infty$, only the very first term is non-zero, and

$$\log M_{Z_n}(t) \longrightarrow \frac{t^2}{2} \qquad \text{so that} \qquad M_{Z_n}(t) \longrightarrow \exp\left\{ \frac{t^2}{2} \right\}$$

INTERPRETATION: Sums of independent and identical distributions have a limiting distribution that is Normal, irrespective of the distribution of the variables.

5.2.2 CONVERGENCE IN DISTRIBUTION

Definition 5.2.1 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 $n = 1, 2, ..$ $F_{X_n}(x)$ =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.

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

$$M_{X_i}(t) \longrightarrow M_X(t)$$

then $X_n \xrightarrow{d} X$

Definition 5.2.2 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

$$f_X(x) = 1 \qquad x = c$$

and zero otherwise, or in other words

$$P[X = c] = 1$$

When the limiting form of the cdf is **not continuous,** as in the degenerate case above, we use a different interpretation of convergence in distribution.

Definition 5.2.3 The sequence of random variables $X_1, ..., X_n$ **converges in distribution** to c if and only if, for all $\epsilon > 0$,

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

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

5.2.3 CONVERGENCE IN PROBABILITY

Definition 5.2.4 CONVERGENCE IN PROBABILITY 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

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

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 5.2.2 *(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 $\epsilon > 0$,

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

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,

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

for all $\epsilon > 0$. Hence

$$P\left[|Y_n - \mu| < \epsilon\right] \longrightarrow 1 \qquad \text{as } n \longrightarrow \infty$$

and $Y_n \xrightarrow{P} \mu$.

Definition 5.2.5 (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 $\epsilon > 0$,

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

Theorem 5.2.3 *For sequence of random variables $X_1, ..., X_n$ if*

$$X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$$

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

5.2.4 ALMOST SURE CONVERGENCE*

Definition 5.2.6 The sequence of random variables $X_1, ..., X_n$ converges almost surely to constant c, denoted

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

if

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

Interpretation : Recall the fundamental definition of a random variable as a real-valued function from sample space Ω to 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 5.2.4 *(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 $\epsilon > 0$,

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

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

Definition 5.2.7 The sequence of random variables $X_1, ..., X_n$ converges almost surely to random variable X, denoted $X_n \stackrel{a.s.}{\longrightarrow} X$, if, for all $\epsilon > 0$,

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

Interpretation: As above, recall that random variable is a real-valued function from sample space Ω to 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 5.2.5 *For sequence of random variables $X_1, ..., X_n$*

$$X_n \stackrel{a.s.}{\longrightarrow} X \Longrightarrow X_n \stackrel{p}{\longrightarrow} X \Longrightarrow X_n \stackrel{d}{\longrightarrow} X$$

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