CPSC 550: Machine Learning II	**2008/9 Term 2**

Lecture 5 — Jan 27, 2009

Lecturer: Nando de Freitas — *Scribe: Gustavo Lacerda*

These notes are based on Wasserman's "All of Statistics" pages 77-82.

Notation

- $\mathbb{V}(X)$ means the variance of X.
- $P_n \rightsquigarrow P$ means "P_n converges in distribution to P".
- $x_n \to x$ means $\lim_{n\to\infty} x_n = x$.
- mgf means "moment-generating function".

The Central Limit Theorem (CLT)

Theorem 5.1 (Central Limit Theorem). *Let $X_1, ..., X_n$ be IID with mean μ and variance σ^2.*
Let $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i$
Let $Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\mathbb{V}(\bar{X}_n)}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$
Then $Z_n \rightsquigarrow Z$, where $Z \sim N(0,1)$.

In other words: for all z, $\lim_{n\to\infty} P(Z_n \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.

That is, for large n, probability statements about $\bar{X}_n$ can be approximated by similar statements about a normal distribution.

Alternative ways to state the CLT

$$Z_n \sim N(0,1) \tag{5.1}$$

$$\bar{X_n} \sim N(\mu, \frac{\sigma^2}{n}) \tag{5.2}$$

$$\bar{X_n} - \mu \sim N(0, \frac{\sigma^2}{n}) \tag{5.3}$$

$$\sqrt{n}(\bar{X_n} - \mu) \sim N(0, \sigma^2) \tag{5.4}$$

$$\frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \sim N(0,1) \tag{5.5}$$

Moment Generating Functions

Let $\mathbb{M}(t)$ be the mgf of a distribution P over random variable X. We now look at the definition and properties of $\mathbb{M}$.

1. Definition:

$$\mathbb{M}(t) = \int e^{tx} p(x) dx$$

   Then the rth moment of $P(X)$ is given by:

$$\mathbb{M}^{(r)}(0) = \mathbb{E}(X^r)$$

   For example:

$$\begin{aligned} \mathbb{M}'(t) &= \int_X e^{tx} p(x) dx \\ \mathbb{M}'(0) &= \int_X p(x) dx = \mathbb{E}(X) \end{aligned}$$

2. If X has mgf $\mathbb{M}_X(t)$ and $Y = aX + b$, then

$$\begin{aligned} \mathbb{M}_Y(t) &= \mathbb{E}(e^{tY}) \\ &= \mathbb{E}(e^{t(aX+b)}) = \mathbb{E}(e^{taX})e^{tb} \\ &= e^{tb}\mathbb{M}_X(ta) \end{aligned}$$

3. Let X, Y be independent random variables with mgfs $\mathbb{M}_X(t)$ and $\mathbb{M}_Y(t)$, let $Z = X + Y$, then

$$\begin{aligned} \mathbb{M}_Z(t) &= \mathbb{E}(e^{tZ}) = \mathbb{E}(e^{tX}e^{tY}) \\ &= \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) = \mathbb{M}_X(t)\mathbb{M}_Y(t) \end{aligned}$$

4. A Guassian example: let $X \sim N(0,1)$,

$$\begin{aligned}
\mathbb{M}_X(t) &= \int_X \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} e^{tx} dx = e^{t^2/2} \\
\mathbb{M}'_X(0) &= 0 \\
\mathbb{M}''_X(0) &= 1
\end{aligned}$$

Lemma 5.2 (based on p.81 (Lemma 5.20) in Wasserman). *The mgf of X is $\mathbb{M}_X(t) = \mathbb{E}(e^{tX})$. Let $\mathbb{M}_n$ be the mgf of Z_n. Let Z be a random variable with mgf $\mathbb{M}$.*

If $\mathbb{M}_n(t) \to \mathbb{M}(t)$ $\forall t$ in an open interval around 0, then $Z_n \rightsquigarrow Z$.

Proof (Central Limit Theorem): Let $Y_i = \dfrac{X_i - \mu}{\sigma}$. Note that since X_i are IID, the Y_i are also IID.

Then $Z_i = n^{-\frac{1}{2}} \sum_i Y_i$.

Let $\mathbb{M}(t)$ be the mgf of Y_1. (Since the Y are IID, we only need one mgf)

The mgf of $\sum_i Y_i$ is

$$\begin{aligned}
\mathbb{E}(e^{t\sum_i X_i}) &= \mathbb{E}(e^{tX_1} \dots e^{tX_n}) && (5.6)\\
&= \mathbb{E}(e^{tX_1}) \dots \mathbb{E}(e^{tX_n}) && (5.7)\\
&= \mathbb{M}(t) \dots \mathbb{M}(t) && (5.8)\\
&= (\mathbb{M}(t))^n && (5.9)
\end{aligned}$$

Since $Z_n = n^{-\frac{1}{2}} \sum_i Y_i$, the mgf of Z_n is

$$\xi_n(t) = \left(\mathbb{M}\left(\frac{t}{\sqrt{n}}\right)\right)^n \tag{5.10}$$

Remember that taking the i-th derivative of the mgf gives the i-th moment. Since the Y_i are standardized in mean and variance, we have:

$$\begin{aligned}
\mathbb{M}'(0) &= \mathbb{E}(Y_i) = 0 && (5.11)\\
\mathbb{M}''(0) &= \mathbb{V}(Y_i) = 1 && (5.12)
\end{aligned}$$

Taylor's theorem guarantees that there is a neighborhood around 0 in which, for all t:

$$\begin{aligned}
\mathbb{M}(t) &= \mathbb{M}(0) + t\mathbb{M}'(0) + \frac{t^2}{2}\mathbb{M}''(0) + \frac{t^3}{3!}\mathbb{M}'''(0) + ... && (5.13)\\
&= 1 + 0 + \frac{t^2}{2} + \frac{t^3}{3!}\mathbb{M}'''(0) + ... && (5.14)\\
&= 1 + \frac{t^2}{2} + \frac{t^3}{3!}\mathbb{M}'''(0) + ... && (5.15)
\end{aligned}$$

$$\begin{aligned}
\text{The mgf of } Z_n \text{ is } \xi_n(t) &= \left(\mathbb{M}\left(\frac{t}{\sqrt{n}}\right)\right)^n && (5.16)\\
&= \left(1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}\mathbb{M}'''(0) + ...\right)^n && (5.17)\\
&= \left(1 + \frac{\frac{t^2}{2} + \frac{t^3}{3!n^{\frac{1}{2}}}\mathbb{M}'''(0) + ...}{n}\right)^n && (5.18)\\
&\to e^{\frac{t^2}{2}} && (5.19)
\end{aligned}$$

The last step follows from the fact that for any sequence $a_n \to a$, it is the case that $\left(1 + \frac{a_n}{n}\right)^n \to e^a$.

This result says that the mgf of Z_n converges to the mgf of a $N(0,1)$ on a neighborhood of 0. Using the lemma, we conclude that Z_n converges in distribution to $Z \sim N(0,1)$. □