Stat 410 Moment Generating Functions

Dr. Scott

September 6, 2005

The random variable results we have discussed follow from useful properties of the Fourier or Laplace transform of the probability density, $f(x)$:

$$\tilde{F}_X(t) = \int f(x)e^{itx}dx = E\left[e^{itX}\right]$$

where $i = \sqrt{-1}$ or

$$M_X(t) = \int f(x)e^{tx}dx = E\left[e^{tX}\right] .$$

These are uniquely defined functions, in 1-1 correspondence with a density. These exist for both discrete and continuous density functions.

Example 1:

Binomial $X \sim B(n,p)$.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \binom{n}{x} p^x \, q^{n-x}$$

letting $q = 1 - p$. The Binomial expansion shows that $f(x)$ sums to 1.

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x \, q^{n-x}$$
$$= (p+q)^n$$
$$= (p + (1-p))^n = 1^n = 1\,.$$

$$
\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= \sum_{x=0}^{n} e^{tx} \cdot \binom{n}{x} p^x \, q^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (e^t \cdot p)^x \, q^{n-x} \\
&= (e^t \cdot p + q)^n
\end{aligned}
$$

also by the Binomial expansion.

Example 2: $Z \sim N(0,1)$.

$$\begin{aligned}
M_Z(t) = E[e^{tZ}] &= \int_{-\infty}^{\infty} e^{tz} f(z) dz \\
&= \int e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2tz+t^2-t^2)} dz \\
&= e^{t^2/2} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} dz \\
&= e^{t^2/2},
\end{aligned}$$

since the integrand is a normal density with $\mu = t$ and $\sigma = 1$.

Example 3: $X \sim N(\mu, \sigma^2)$

Equivalently, $X = \mu + \sigma Z$. Then

$$\begin{aligned}
M_X(t) &= E[e^{tX}] \\
&= E[e^{t(\mu+\sigma Z)}] \\
&= E[e^{t\mu+t\sigma Z}] = E[e^{t\mu}\, e^{t\sigma Z}] \\
&= e^{t\mu}\, E[e^{t\sigma Z}] \\
&= e^{t\mu}\, M_Z(t\sigma) \\
&= e^{t\mu}\, e^{(t\sigma)^2/2} \\
&= \text{exp}\{\mu t + \frac{1}{2}\sigma^2 t^2\}\,.
\end{aligned}$$

Example 4: $Y = \sum_{i=1}^{n} X_i$ where X_i is a random sample from $N(\mu, \sigma^2)$.

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t(X_1+X_2+\cdots+X_n)}] \\
&= E[e^{tX_1}\, e^{tX_2} \cdots e^{tX_n}] \\
&= E[e^{tX_1}]\, E[e^{tX_2}] \cdots E[e^{tX_n}] \\
&= E[e^{tX_1}]^n \\
&= \exp\{\mu t + \sigma^2 t^2/2\}^n \\
&= \exp\{n\mu t + n\sigma^2 t^2/2\} \\
&\sim N(n\mu, n\sigma^2)\,.
\end{aligned}
$$

It follows that $\bar{X} \sim N(\mu, \sigma^2/n)$.

Example 5: Relevant for our regression problem:

$$S = \sum_{i=1}^{n} w_i Y_i \quad \text{where} \quad Y_i \sim N(\mu_i, \sigma_i^2).$$

$$\begin{aligned} M_S(t) &= E[e^{tS}] \\ &= E[e^{tw_1 Y_1}\, e^{tw_2 Y_2} \cdots e^{tw_n Y_n}] \\ &= E[e^{tw_1 Y_1}]\, E[e^{tw_2 Y_2}] \cdots E[e^{tw_n Y_n}] \\ &= \prod_{i=1}^{n} M_{Y_i}(w_i t) \\ &= \prod_{i=1}^{n} \exp\left[\mu_i (w_i t) + \sigma_i^2 (w_i t)^2/2\right] \end{aligned}$$

Continuing

$$M_S(t) = \prod_{i=1}^{n} \exp\left[\mu_i(w_i t) + \sigma_i^2 (w_i t)^2/2\right]$$
$$= \exp\left[\mu t + \sigma^2 t^2 2\right]$$

where

$$\mu = \sum_{i=1}^{n} w_i \mu_i \qquad \text{and}$$
$$\sigma^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 \, ,$$

and $S \sim N(\mu, \sigma^2)$.

Section 2.1 Inferences in Regression

$$\hat{\beta}_1 = b_1 = \frac{\sum(x_i - \bar{x})(Y_i - \bar{Y})}{\sum(x_i - \bar{x})^2}$$

$$= \frac{\sum(x_i - \bar{x})Y_i}{\sum(x_i - \bar{x})^2} = \sum w_i Y_i$$

since $\sum(x_i - \bar{x})\bar{Y} = 0$ where

$$w_i = \frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}$$

Note that

$$\sum w_i = \frac{1}{\sum(x_i - \bar{x})^2} \sum(x_i - \bar{x}) = 0$$

$$\sum w_i^2 = \frac{1}{[\sum(x_i - \bar{x})^2]^2} \sum(x_i - \bar{x})^2$$
$$= \frac{1}{\sum(x_i - \bar{x})^2}$$

and

$$\sum w_i x_i = \frac{1}{\sum(x_i - \bar{x})^2} \sum(x_i^2 - \bar{x} x_i)$$
$$= \frac{1}{\sum x_i^2 - n\bar{x}^2} \left(\sum x_i^2 - n\bar{x}^2\right)$$
$$= 1\,.$$

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \sim N(\beta_0 + \beta_1 x_i, \sigma_\epsilon^2)$$

so $\mu_i = \beta_0 + \beta_1 x_i$ and $\sigma_i^2 = \sigma_\epsilon^2$.

Thus b_1 is normal with moments

$$\begin{aligned} E[b_1] &= \sum w_i \mu_i \\ &= \sum w_i(\beta_0 + \beta_1 x_i) \\ &= \beta_0 \sum w_i + \beta_1 \sum w_i x_i \\ &= \beta_0 \cdot 0 + \beta_1 \cdot 1 = \beta_1 \end{aligned}$$

$$\begin{aligned} Var[b_1] &= \sum w_i^2 \sigma_i^2 \\ &= \sigma_\epsilon^2 / \sum (x_i - \bar{x})^2 \end{aligned}$$

$$b_1 \sim N\left(\beta_1, \frac{\sigma_\epsilon^2}{\sum(x_i - \bar{x})^2}\right)$$

Thus b_1 is unbiased, consistent, and

$$s^2(b_1) = \frac{\text{MSE}}{\sum(x_i - \bar{x})^2}$$

so

$$\frac{b_1 - \beta_1}{s(b_1)} \sim t_{n-2} \quad \text{is the pivot}$$

for finding a confidence interval for β_1 or testing $H_0 : \beta_1 = 0$.

Read Section 2.1 for details and examples.

A similar derivation for $\widehat{\beta}_0 = b_0$ shows

$$b_0 \sim N\left(\beta_0, \sigma_\epsilon^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right]\right) .$$

With

$$s^2(b_0) = \text{MSE}\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}\right]$$

the pivot for the parameter β_0 is

$$\frac{b_0 - \beta_0}{s(b_0)} \sim t_{n-2}$$

see Section 2.2. Read Section 2.3.

Sec 2.4: At a new point, x_h, the prediction

$$\widehat{Y}_h = b_0 + b_1 x_h$$

is a linear combination of normals, so normal

$$\widehat{Y}_h \sim N\left(\beta_0 + \beta_1 x_h, \sigma_\epsilon^2 \left[\frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right]\right)$$

so replacing σ_ϵ^2 with MSE, the pivot is

$$\frac{\widehat{Y}_h - (\beta_0 + \beta_1 x_h)}{s(\widehat{Y}_h)} \sim t_{n-2}$$

(Note: Choose $x_h = 0$ and $\widehat{Y}_h \equiv b_0$.)

Section 2.5 discusses predicting a range of values for Y_h at $x = x_h$ rather than just its (conditional) average, $\widehat{Y}_h$.

Whereas the central limit theorem suggests normality even if the noise, ϵ_i is not normal, the assumption of normality is key here. It turns out that

$$s^2(Y_h) = \text{MSE} + s^2(\widehat{Y}_h)$$

leading to prediction interval (2.36).

Section 2.6. Many prediction intervals.

With each test or confidence interval, we are allowing ourselves a 5% chance for error. If we make two such statements, is our chance of making a mistake still 5%, or is it greater?

The answer is that it is greater.

$$Prob(no\ error) = .95 \times .95 = .9025$$

Ouch! Our overall significance level is 9.75%.

With 10 tests, $\alpha = 40.13\%$ overall!

Working and Hotelling showed that for the special case of linear regression, an overall 5% confidence band for the *entire* regression line could be guaranteed if we replace the t_{n-2} pivotal quantity with

$$W = \sqrt{2\,F_{1-\alpha}(2, n-2)}$$

Which would be more appropriate for the homework? You can use either one this time.

Read the rest of Section 2.6.