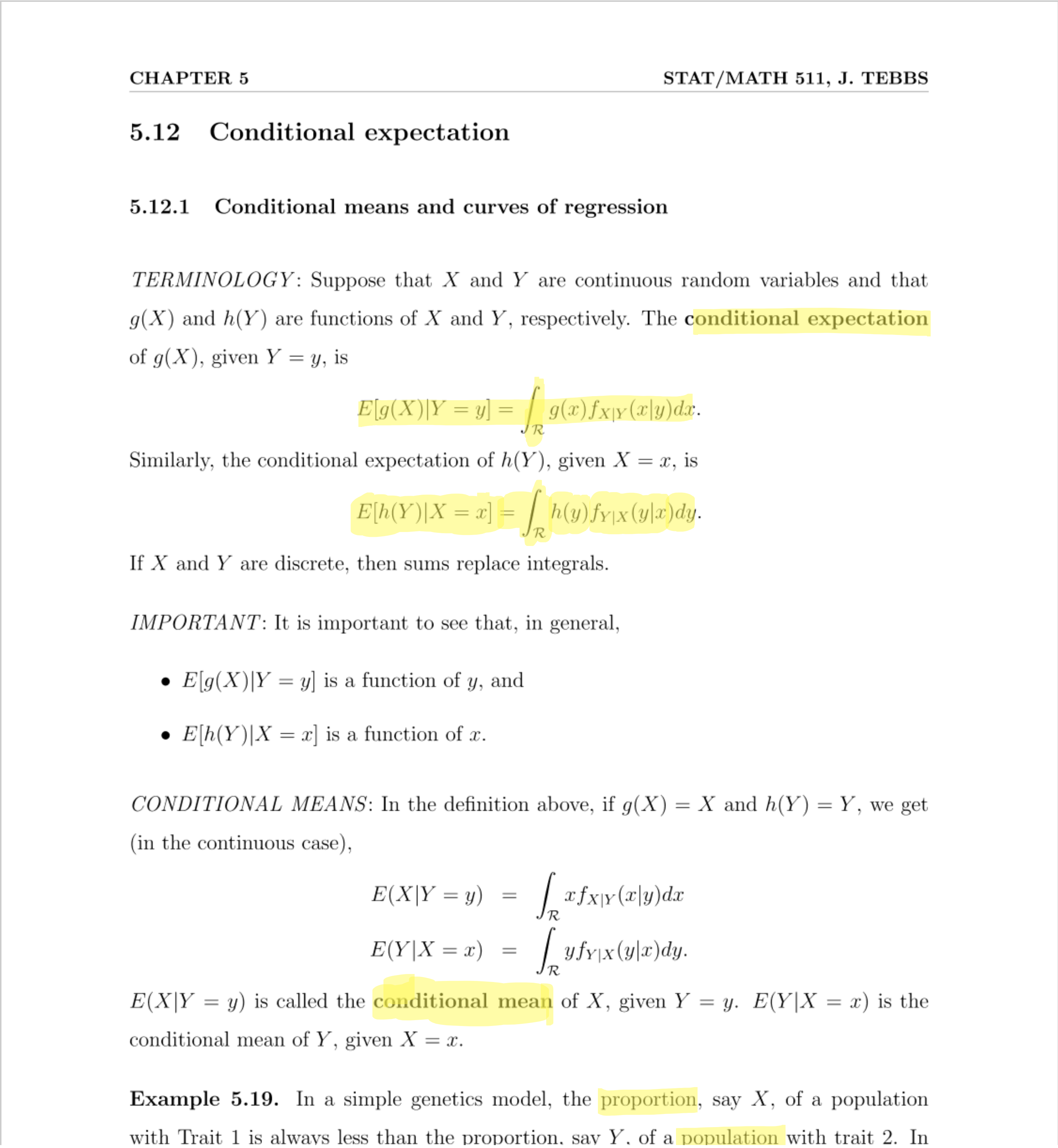
Section 5.12

Tuesday, November 15, 2016 12:35 PM

Section 5.12



5.12 Conditional expectation

5.12.1 Conditional means and curves of regression

TERMINOLOGY: Suppose that X and Y are continuous random variables and that $g(X)$ and $h(Y)$ are functions of X and Y, respectively. The **conditional expectation** of $g(X)$, given $Y = y$, is

$$E[g(X)|Y = y] = \int_{\mathcal{R}} g(x) f_{X|Y}(x|y)dx.$$

Similarly, the conditional expectation of $h(Y)$, given $X = x$, is

$$E[h(Y)|X = x] = \int_{\mathcal{R}} h(y) f_{Y|X}(y|x)dy.$$

If X and Y are discrete, then sums replace integrals.

IMPORTANT: It is important to see that, in general,

- $E[g(X)|Y = y]$ is a function of y, and
- $E[h(Y)|X = x]$ is a function of x.

CONDITIONAL MEANS: In the definition above, if $g(X) = X$ and $h(Y) = Y$, we get (in the continuous case),

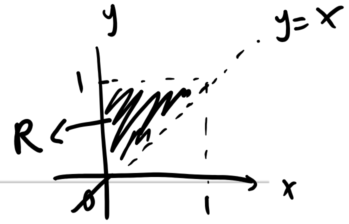
$$\begin{aligned} E(X|Y = y) &= \int_{\mathcal{R}} x f_{X|Y}(x|y)dx \\ E(Y|X = x) &= \int_{\mathcal{R}} y f_{Y|X}(y|x)dy. \end{aligned}$$

$E(X|Y = y)$ is called the **conditional mean** of X, given $Y = y$. $E(Y|X = x)$ is the conditional mean of Y, given $X = x$.

Example 5.19. In a simple genetics model, the proportion, say X, of a population with Trait 1 is always less than the proportion, say Y, of a population with trait 2. In

Example 5.19. In a simple genetics model, the proportion, say X, of a population with Trait 1 is always less than the proportion, say Y, of a population with trait 2. In Example 5.3, we saw that the random vector (X, Y) has joint pdf

$$f_{X,Y}(x,y) = \begin{cases} 6x, & 0 < x < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$



$X \sim f_X(x) \quad E(X) = \int x f_X(x)\,dx$

In Example 5.5, we derived the conditional distributions

$$f_{X|Y}(x|y) = \begin{cases} 2x/y^2, & 0 < x < y \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x}, & x < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the conditional mean of X, given $Y = y$ is

$$\begin{aligned} E(X|Y=y) &= \int_0^y x f_{X|Y}(x|y)dx \\ &= \int_0^y x\left(\frac{2x}{y^2}\right)dx = \frac{2}{y^2}\left(\frac{x^3}{3}\bigg|_0^y\right) = \frac{2y}{3}. \end{aligned}$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Similarly, the conditional mean of Y, given $X = x$ is

$$\begin{aligned} E(Y|X=x) &= \int_x^1 y f_{Y|X}(y|x)dy \\ &= \int_x^1 y\left(\frac{1}{1-x}\right)dy = \frac{1}{1-x}\left(\frac{y^2}{2}\bigg|_x^1\right) = \frac{1}{2}(x+1). \end{aligned}$$

That $E(Y|X=x) = \frac{1}{2}(x+1)$ is not surprising because $Y|\{X=x\} \sim \mathcal{U}(x,1)$. ☐

TERMINOLOGY: Suppose that (X, Y) is a bivariate random vector.

- The graph of $E(X|Y=y)$ versus y is called the **curve of regression** of X on Y.
- The graph of $E(Y|X=x)$ versus x is the curve of regression of Y on X.

The curve of regression of Y on X, from Example 5.19, is depicted in Figure 5.17.

5.12.2 Iterated means and variances

5.12.2 Iterated means and variances

REMARK: In general, $E(X|Y=y)$ is a function of y, and y is fixed (not random). Thus, $E(X|Y=y)$ is a **fixed number**. However, $E(X|Y)$ is a function of Y; thus, $E(X|Y)$ is a **random variable**! Furthermore, as with any random variable, it has a mean and variance associated with it!!

ITERATED LAWS: Suppose that X and Y are random variables. Then the **laws of iterated expectation and variance**, respectively, are given by

$$E(X) = E[E(X|Y)]$$

$$\underset{\substack{\downarrow \\ f_Y}}{E}\Big[\underset{\substack{\downarrow \\ f_{X|Y}}}{E}(X|Y)\Big]$$

$$(X,Y) \sim f_{X,Y}(x,y)$$

$$E(X) = \iint x f_{X,Y}(x,y)\,dx\,dy$$

$$= \int \frac{\int x f_{X,Y}(x,y)\,dx}{f_Y(y)} \cdot f_Y(y)\,dy = \int \int x \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)}\,dx\; f_Y(y)\,dy$$

$$= \int \left[\int x f_{X|Y}(x|y)\,dx\right] f_Y(y)\,dy$$

$$= \int E[X|Y] \cdot f_Y(y)\,dy$$

$$= E[E(X|Y)]$$

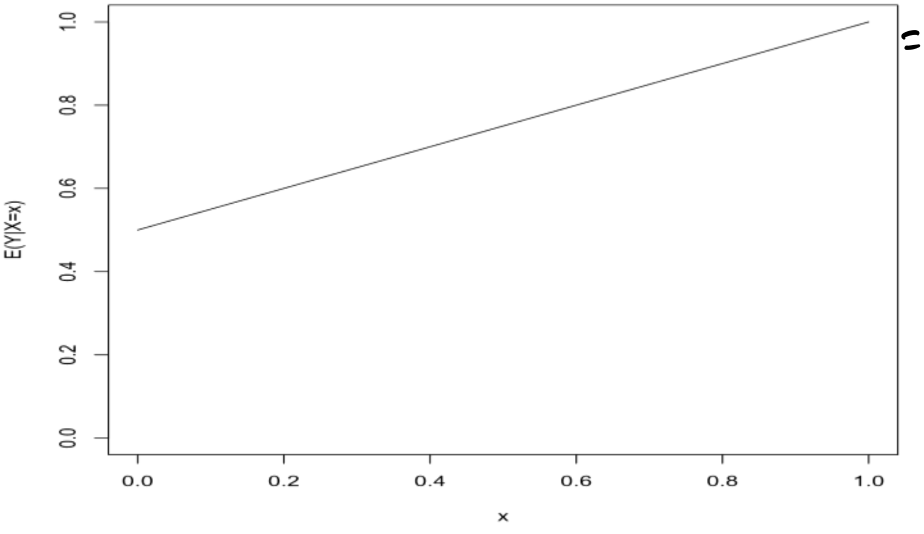


Figure 5.17: *The curve of regression* $E(Y|X=x)$ *versus* x *in Example* 5.19.

and

$$V(X) = E[V(X|Y)] + V[E(X|Y)].$$

$$V(X) = E[V(X|Y)] + V[E(X|Y)].$$

NOTE: When considering the quantity $E[E(X|Y)]$, the inner expectation is taken with respect to the conditional distribution $f_{X|Y}(x|y)$. However, since $E(X|Y)$ is a function of Y, the outer expectation is taken with respect to the marginal distribution $f_Y(y)$.

Proof. We will prove that $E(X) = E[E(X|Y)]$ for the continuous case. Note that

$$\begin{aligned} E(X) &= \int_{\mathcal{R}}\int_{\mathcal{R}} x f_{X,Y}(x,y)dxdy \\ &= \int_{\mathcal{R}}\int_{\mathcal{R}} x f_{X|Y}(x|y) f_Y(y)dxdy \\ &= \int_{\mathcal{R}} \underbrace{\left[\int_{\mathcal{R}} x f_{X|Y}(x|y)dx\right]}_{E(X|Y=y)} f_Y(y)dy = E[E(X|Y)]. \quad \square \end{aligned}$$

Example 5.20. Suppose that in a field experiment, we observe Y, the number of plots, out of n, that respond to a treatment. However, we don't know the value of p, the probability of response, and furthermore, we think that it may be a function of location,

temperature, precipitation, etc. In this situation, it might be appropriate to regard p as a random variable. Specifically, suppose that the random variable P varies according to a beta(α, β) distribution. That is, we assume a **hierarchical structure**:

$$\begin{aligned} Y|P = p &\sim \text{binomial}(n,p) \\ P &\sim \text{beta}(\alpha,\beta). \end{aligned}$$

The (unconditional) mean of Y can be computed using the iterated expectation rule:

$$E(Y) = E[E(Y|P)] = E[nP] = nE(P) = n\left(\frac{\alpha}{\alpha+\beta}\right).$$

The (unconditional) variance of Y is given by

$$\begin{aligned} V(Y) &= E[V(Y|P)] + V[E(Y|P)] \\ &= E[nP(1-P)] + V[nP] \end{aligned}$$

$$
\begin{aligned}
V(Y) &= E[V(Y|P)] + V[E(Y|P)] \\
&= E[nP(1-P)] + V[nP] \\
&= nE(P - P^2) + n^2V(P) \\
&= nE(P) - n\{V(P) + [E(P)]^2\} + n^2V(P) \\
&= n\left(\frac{\alpha}{\alpha+\beta}\right) - n\left[\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} + \left(\frac{\alpha}{\alpha+\beta}\right)^2\right] + \frac{n^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= n\left(\frac{\alpha}{\alpha+\beta}\right)\left[1 - \left(\frac{\alpha}{\alpha+\beta}\right)\right] + \underbrace{\frac{n(n-1)\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}_{\text{extra variation}}.
\end{aligned}
$$

Unconditionally, the random variable Y follows a **beta-binomial** distribution. This is a popular probability model for situations wherein one observes binomial type responses but where the variance is suspected to be larger than the usual binomial variance. ☐

BETA-BINOMIAL PMF: The probability mass function for a **beta-binomial** random variable Y is given by

$$
\begin{aligned}
p_Y(y) = \int_0^1 f_{Y,P}(y,p)dp &= \int_0^1 f_{Y|P}(y|p)f_P(p)dp \\
&= \int_0^1 \binom{n}{y} p^y(1-p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp \\
&= \binom{n}{y} \frac{\Gamma(\alpha+\beta)\Gamma(y+\alpha)\Gamma(n+\beta-y)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)},
\end{aligned}
$$

for $y = 0, 1, ..., n$, and $p_Y(y) = 0$, otherwise.