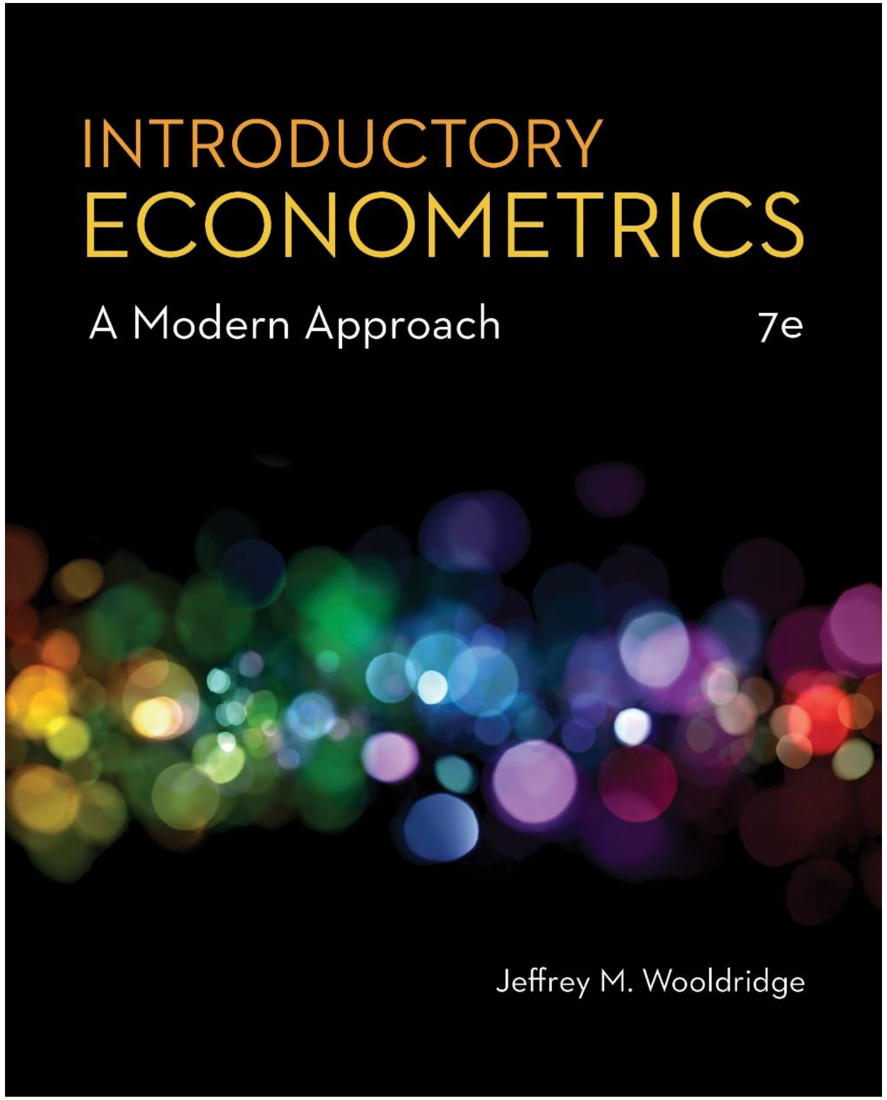
Chapter 17

Limited Dependent Variable Models and Sample Selection



Limited Dependent Variable Models

- Broadly defined as a dependent variable whose range of values is substantively restricted

- Dependent variable is qualitative (qualitative response or discrete choice models) or limited in their range

Limited Dependent Variable Models (cont)

- **Binary variable** takes on only two values, zero and one

- Generally discrete response variables - y takes on **a small number of integer values** (the number of times a young man is arrested during a year, or the number of children born to a woman or **“choice” between multiple**, more than two outcomes – being full time/part time employed or unemployed, payment method or travel mode choice)

Limited Dependent Variable Models (cont)

- Truncation occurs when sample data are drawn from **a subset of a larger population of interest** (studies of income based on incomes above or below some poverty line may be of limited usefulness for inference about the whole population)

- Censoring - suppose that instead of being unobserved, all incomes below the poverty line are reported as if they were at the poverty line (introduces a **distortion into conventional statistical results** that is like that of truncation)

- However, censoring is essentially **a defect in the sample data**

Multiple Regression Analysis

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + u$
- Dummy Variables (Ch 7)

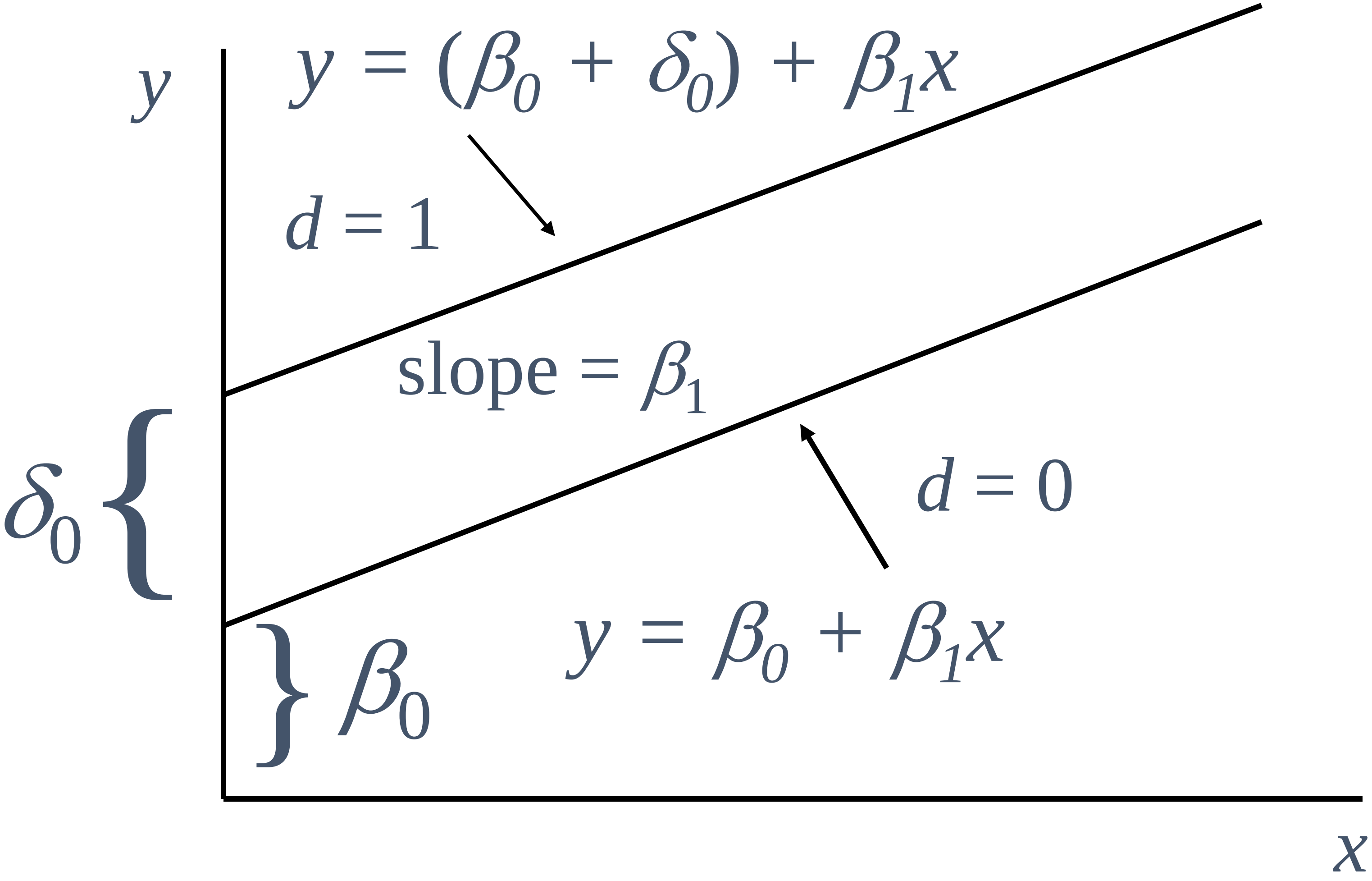
Dummy Variables

- A dummy variable is a variable that takes on the value 1 or 0
- Examples: male (= 1 if are male, 0 otherwise), south (= 1 if in the south, 0 otherwise), etc.
- Dummy variables are also called binary variables, for obvious reasons
- Dummy variable trap (number of categories minus one)

A Dummy Independent Variable

- Consider a simple model with one continuous variable (*x*) and one dummy (*d*)
- $y = \beta_0 + \delta_0 d + \beta_1 x + u$
- This can be interpreted as **an intercept shift (and/or slope shift)**
- If d = 0, then $y = \beta_0 + \beta_1 x + u$
- If d = 1, then $y = (\beta_0 + \delta_0) + \beta_1 x + u$
- The case of d = 0 is the **base group** or **benchmark group**

Example of $\delta_0 > 0$



Linear Probability Model

- $P(y = 1|x) = E(y/x)$, when y is a binary variable, so we can write our model as:

$$P(y = 1|x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

- So, the interpretation of β_j **is the change in the probability** of success when x_j changes
- The predicted y is the predicted probability of success or outcome 1 (Y=1)
- Potential problem that can be outside [0,1]

Linear Probability Model (cont)

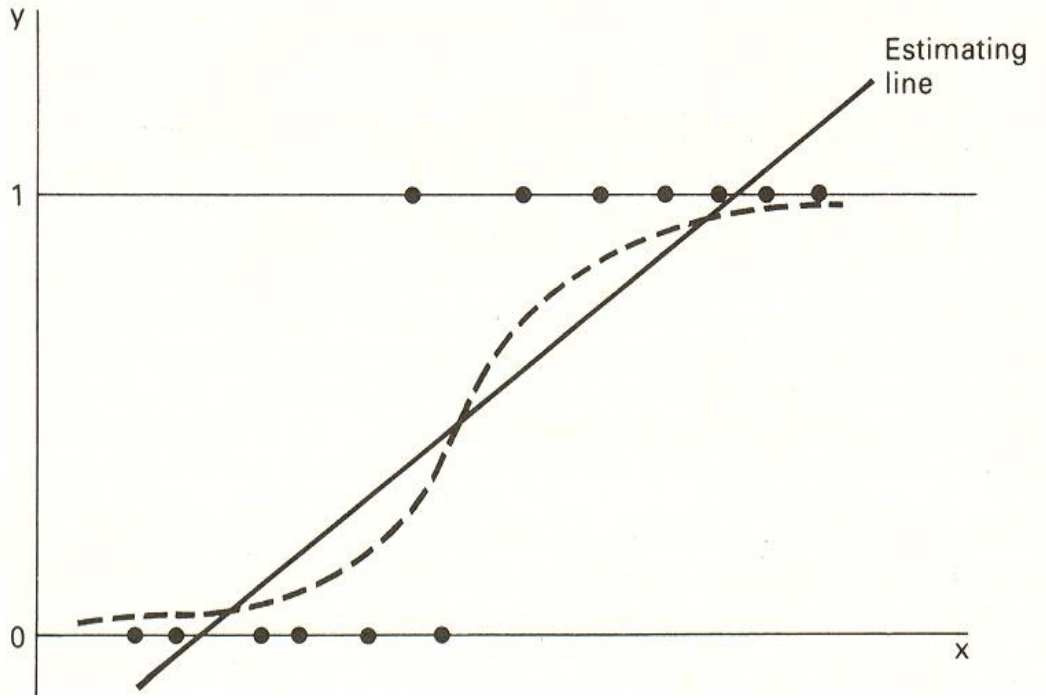
- Even without predictions outside of [0,1], we may estimate effects that imply a change in *x* changes the probability by more than +1 or – 1, so best to use changes near mean

- This model will violate assumption of homoskedasticity, so will affect inference

- LPM produces both **nonsense probabilities and negative variances**

- Despite drawbacks, it's usually a good place to start when *y* is binary

LPM and nonlinear specifications



- Shortcomings of the LPM
- **Major flaw**: ***linear change*** in the probability that Y=1 associated with a *unit change* in X
- As regressions with a binary dependent variable Y models the probability that Y=1, it make sense **to adopt a nonlinear formulation** (predicted values are forced to be between 0-1) – **prob. is S- shaped function of X**

Application1: Bank decision on the mortgage applications

- Source: J. Stock and M. Watson, *Introduction to Econometrics*, Addison Wesley, Pearson International Edition, 2003

- HMDA data (*Home Mortgage Disclosure Act*) - Cross-sectional data, mortgage applications made in 1990 in the greater Boston metropolitan area using a subset of the original dataset (N=2380)

Application1: Bank decision on the mortgage applications (cont)

TABLE 9.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
Financial Variables		
P/I ratio	Ratio of total monthly debt payments to total monthly income	0.331
housing expense-to-income ratio	Ratio of monthly housing expenses to total monthly income	0.255
loan-to-value ratio	Ratio of size of loan to assessed value of property	0.738
consumer credit score	1 if no "slow" payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
mortgage credit score	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
public bad credit record	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074

TABLE 9.1 Variables Included in Regression Models of Mortgage Decisions

Variable	Definition	Sample Average
Additional Applicant Characteristics		
denied mortgage insurance	1 if applicant applied for mortgage insurance and was denied, 0 otherwise	0.020
self-employed	1 if self-employed, 0 otherwise	0.116
single	1 if applicant reported being single, 0 otherwise	0.393
high school diploma	1 if applicant graduated from high school, 0 otherwise	0.984
unemployment rate	1989 Massachusetts unemployment rate in the applicant's industry	3.8
condominium	1 if unit is a condominium, 0 otherwise	0.288
black	1 if applicant is black, 0 if white	0.142
deny	1 if mortgage application denied, 0 otherwise	0.120

FIGURE 9.1 Scatterplot of Mortgage Application Denial and the Payment-to-Income Ratio

Mortgage applicants with a high ratio of debt payments to income (*P/I ratio*) are more likely to have their application denied (*deny* = 1 if denied, *deny* = 0 if approved). The linear probability model uses a straight line to model the probability of denial, conditional on the *P/I ratio*.

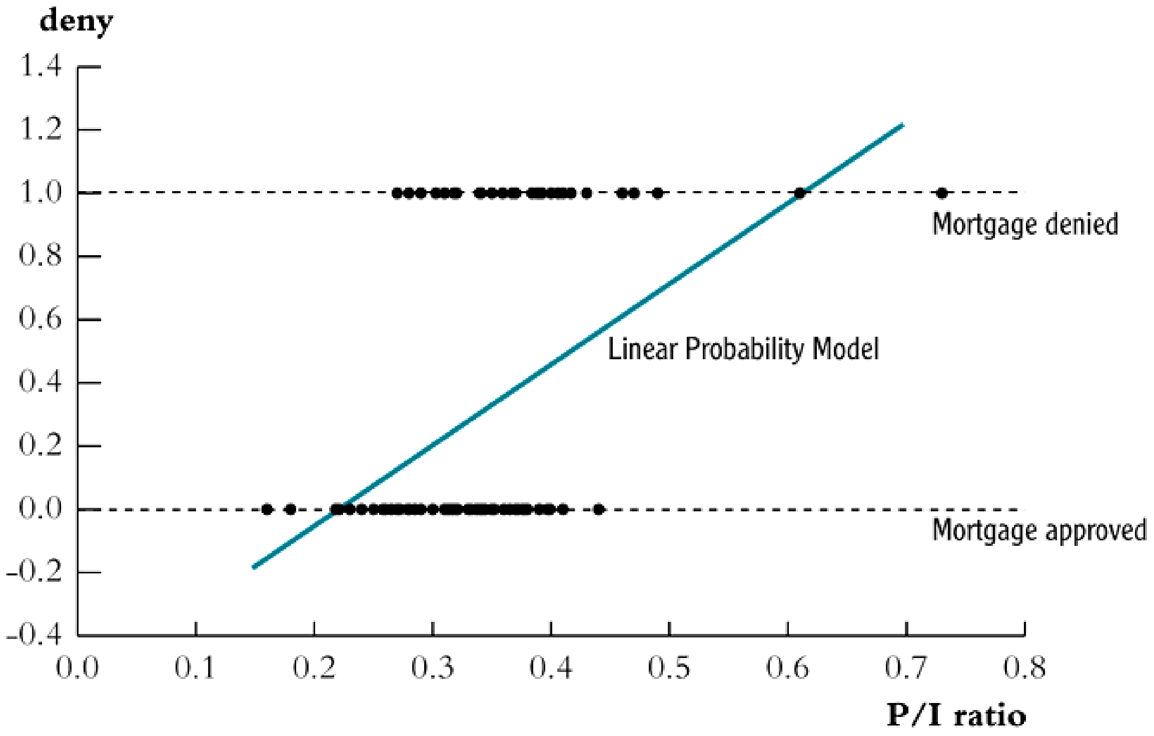
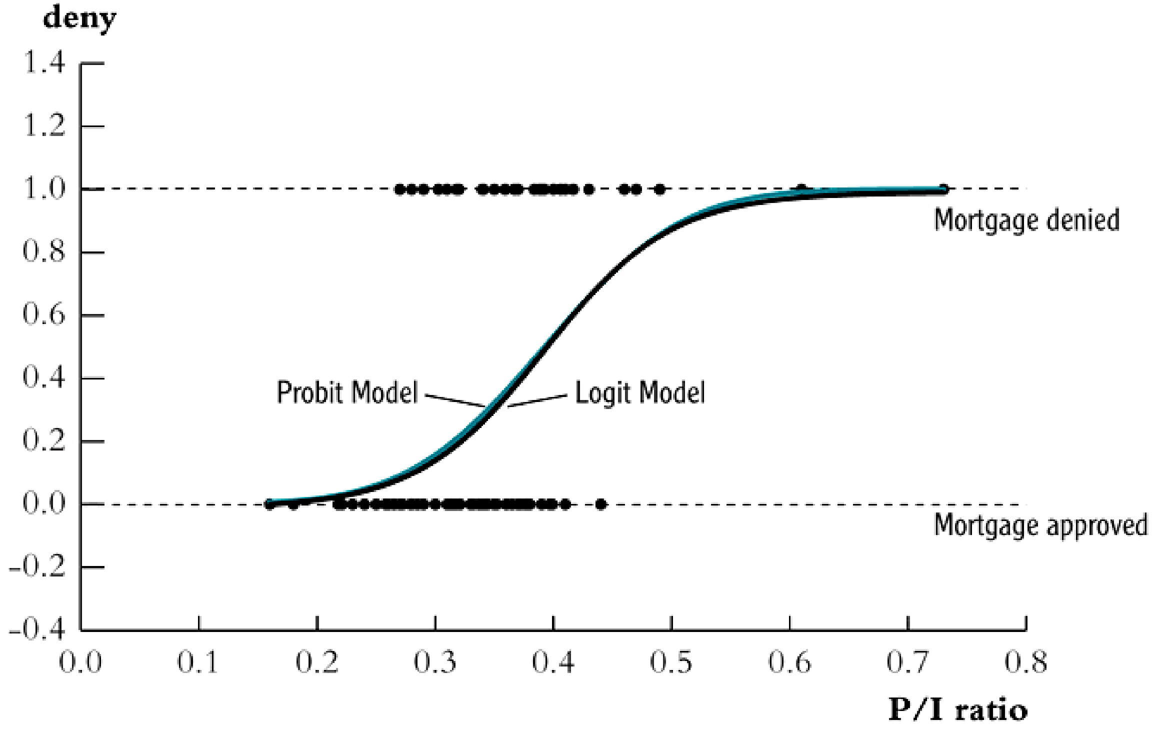


FIGURE 9.3 **Probit and Logit Models of the Probability of Denial, Given the P/I Ratio**

These logit and probit models produce nearly identical estimates of the probability that a mortgage application will be denied, given the payment-to-income ratio.



Limited Dependent Variables

- $P(y = 1|\boldsymbol{x}) = G(\beta_0 + \boldsymbol{x\beta})$, where G is a function: $0 < G(z) < 1$

- Various nonlinear functions have been suggested for the function G

- The two are used in the vast majority of applications (*logit* and *probit*)

The Logit Model

- One common choice for G(z) is the logistic function, which is the *cdf* for a **standard logistic random variable**

- $G(z)$ = p = exp(z)/[1 + exp(z)] = $\Lambda(z)$,

where z is linear function of the explanatory variables.

- This case is referred to as a **logit model**, or sometimes as **a logistic regression** (odds ratio of p and (1-p)): $\left[\frac{p}{(1-p)}\right] = e^{z} = e^{X\beta}$

The Probit Model

- Another choice for $G(z)$ is the standard normal cumulative distribution function (*cdf*)

- $G(z) = \int g(z)dz$, where $g(z)$ is the standard normal, so $g(z) = (2\pi)^{-1/2}\exp(-z^2/2)$, or:

$$\Pr o\, b(Y) = 1) = \int_{-\infty}^{\beta\prime x} \frac{1}{\sqrt{2\pi}} \exp[-(z^2/2)]\, dz = \int_{-\infty}^{\beta\prime x} g(z)\, dz = F(z) = \Phi(z)$$

-

- This case is referred to as a probit model

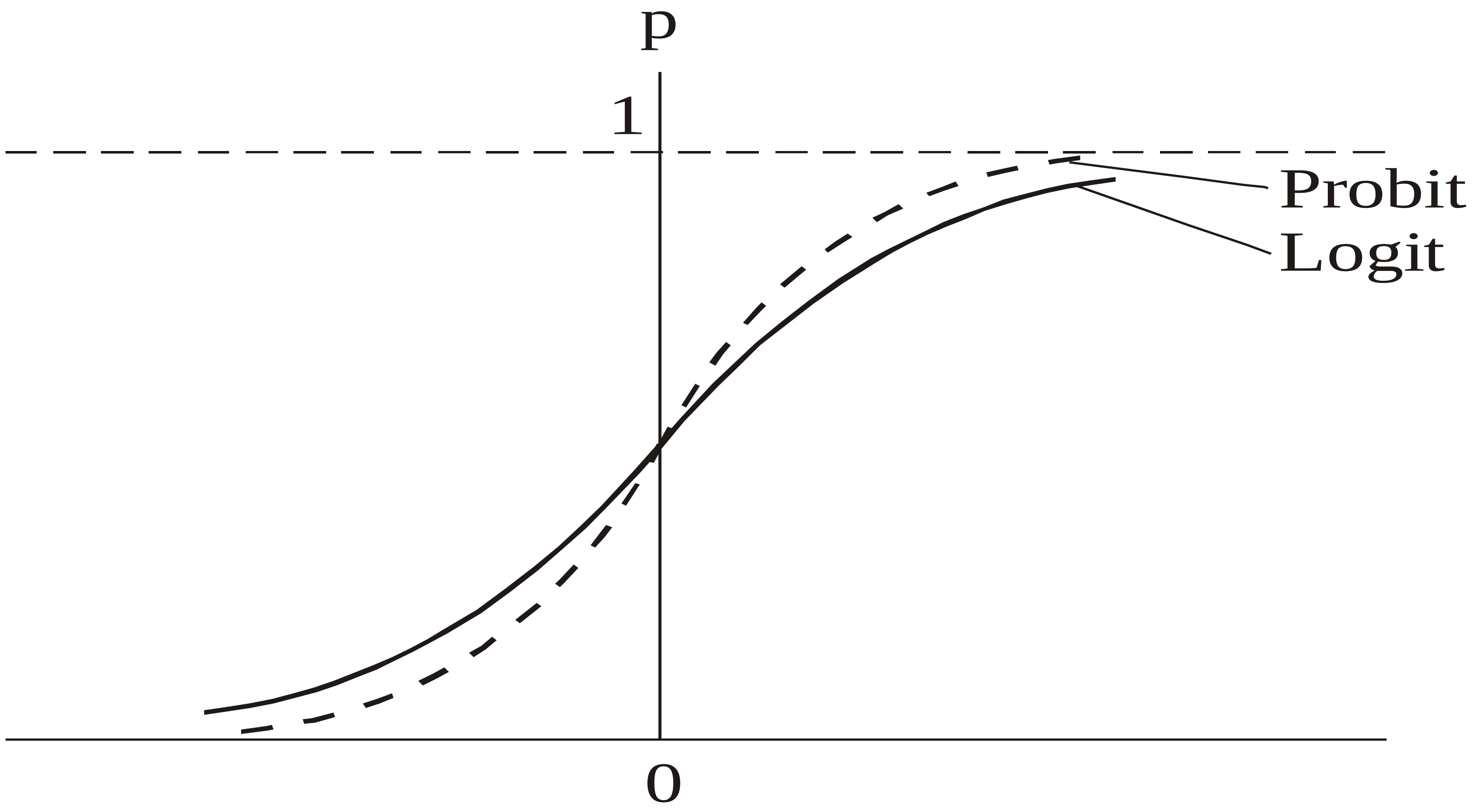
Logit and Probit Model

- Both functions have **similar shapes** – they are increasing in *z*, most quickly around 0

- Since it is a nonlinear model, it cannot be estimated by our usual methods

- Use **maximum likelihood estimation (MLE)**

Two cumulative distribution functions



Alternative definition of logit and probit model

- Two nonlinear models can be derived from an **underlying latent variable model**

- Let y* be an unobserved, or latent, variable, and suppose that:

$$y^* = \beta_0 + \boldsymbol{x\beta} + \boldsymbol{e},\ y = 1\,[y^* > 0]$$

  where we introduce the notation $1[.]$to define a binary outcome

- The function $1[.]$ is called **the indicator function**, which takes on the value one if the event in brackets is true, and zero otherwise.

Latent Variables

- The idea is that there is an underlying variable y*, that can be modeled as:

$$y^* = \beta_0 + \boldsymbol{x\beta} + e,$$

but we only observe $y = 1$, if $y^* > 0$, and $y = 0$ if $y^* \leq 0$

- We assume that **e is independent of x** and that **e** has either the standard logistic distribution or the standard normal distribution (logit or probit model)

- In other case, **e** is symmetrically distributed around zero, which means that 1-G(-z)=G(z) for all real number z

- The latent variable formulation tends to give the impression that we are primarily interested in the effects of each x_j on y*

Binary Dependent Variable Models (summary)

- LPM: $Pr\,(Y=1 \mid X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$
- Probit Model: $Pr\,(Y=1 \mid X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k)$
- Logit Model: $Pr\,(Y=1 \mid X) = \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k) =$

$$= \frac{1}{1 + e^{-\,(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k)}}$$

or

Logistic regression:

$$\left[\frac{p}{(1-p)}\right] = e^{X\beta}$$

Probits and Logits

- Both the probit and logit are nonlinear and **require maximum likelihood estimation** (more on fundamentals is forthcoming)
- No real reason to prefer one over the other
- Traditionally saw more of the logit, mainly because the logistic function leads to a more easily computed model
- Today, probit is easy to compute with standard packages, so more popular

Interpretation of Probit and Logit (vs LPM)

- In general, we care about the effect of x on P(y = 1/$\boldsymbol{x}$), that is, we care about $\partial p/\partial x$

- For the linear case, this is easily computed as the coefficient on x

- To find the partial effect of roughly continuous variables on the response probability, we must rely on calculus

- For the nonlinear probit and logit models, it's more complicated:

$$\partial p/\partial x_j = g(\beta_0 + x\beta)\,\beta_j, \text{ where } g(z) = dG/dz$$

- e.g., if G(z) is the cumulative standardized normal distribution (cdf), its derivative, is just the standardized normal distribution itself (pdf)

BINARY CHOICE MODELS: LOGIT ANALYSIS

$$p = G(Z) = \Lambda(Z) = \frac{1}{1+e^{-Z}}$$

$$p_i = G(Z_i) = \frac{1}{1+e^{-\beta_0-\beta_1 X_i}}$$

$$\frac{\partial G}{\partial X} = \frac{dG}{dZ}\frac{\partial Z}{\partial X} = g(Z)\beta_1 = \frac{e^{-Z}}{(1+e^{-Z})^2}\beta_1$$

$$g(Z) = \frac{\mathrm{d}G}{\mathrm{d}Z} = \frac{e^{-Z}}{(1+e^{-Z})^2}$$

To calculate marginal effect **of X on G (z) we can calculate the differential directly, but also by breaking it up into two stages (*G* is function of Z, and Z is a function of X) – very useful if Z is a function of more than one variable. Marginal effect varies with X**

BINARY CHOICE MODELS: LOGIT ANALYSIS

$$p = G(Z) = \frac{1}{1+e^{-Z}}$$

$$Y = \frac{U}{V}$$

$$\frac{dY}{dZ} = \frac{V\frac{dU}{dZ} - U\frac{dV}{dZ}}{V^2}$$

$$U = 1 \quad \Rightarrow \quad \frac{\mathrm{d}U}{\mathrm{d}Z} = 0$$

$$V = 1 + e^{-Z} \quad \Rightarrow \quad \frac{\mathrm{d}V}{\mathrm{d}Z} = -e^{-Z}$$

$$\frac{\mathrm{d}p}{\mathrm{d}Z} = \frac{\left(1+e^{-Z}\right)\times 0 - 1\times\left(-e^{-Z}\right)}{\left(1+e^{-Z}\right)^2} = \frac{e^{-Z}}{\left(1+e^{-Z}\right)^2}$$

We apply the rule to the expression for *G(Z)*/p

BINARY CHOICE MODELS: LOGIT ANALYSIS

$$p = G(Z) = \frac{1}{1 + e^{-Z}}$$

$$Z = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k$$

$$\frac{\partial G}{\partial X_j} = \frac{\mathrm{d}G}{\mathrm{d}Z} \frac{\partial Z}{\partial X_j} = g(Z)\beta_j = \frac{e^{-Z}}{(1 + e^{-Z})^2} \beta_j$$

$$g(Z) = \frac{\mathrm{d}G}{\mathrm{d}Z} = \frac{e^{-Z}}{(1 + e^{-Z})^2}$$

We will do this theoretically for the general case where *Z* is a function of several explanatory variables. The marginal effect is not constant because it depends on the value of *Z*, which in turn depends on the values of the explanatory variables. A common procedure is to evaluate it for the sample means of the explanatory variables.

BINARY CHOICE MODELS: PROBIT ANALYSIS

$$p = G(Z)$$

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k$$

$$\frac{\partial G}{\partial X_j} = \frac{\mathrm{d}G}{\mathrm{d}Z}\frac{\partial Z}{\partial X_j} = g(Z)\beta_j = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}\right)\beta_j$$

$$g(Z) = \frac{\mathrm{d}G}{\mathrm{d}Z} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$$

The marginal effect of X_j on G (p) can be written as the product of the marginal effect of Z on G and the marginal effect of X_j on Z. The marginal effect of Z on G is given by the standardized normal distribution. The marginal effect of X_j on Z is given by β_j.

As with logit analysis, the marginal effects vary with Z. A common procedure is to evaluate them for the value of Z given by the sample means of the explanatory variables.

Interpretation (continued)

- Clear that it's incorrect to just compare the coefficients across the three models

- Can compare sign and significance (based on a standard ***z*** test) of coefficients, though

- To compare the magnitude of effects, need to calculate the partial derivatives, **say at the means (partial effect at the average, PEA or average marginal effect, AME)**

- Stata will do this for you in the probit cases (both cases)

Marginal effect of binary/discrete explanatory variable

- The partial effect from change x_1 from zero to one:

$$G(\beta_0 + \beta_1 * 1 + \beta_2 X_2 + \ldots \beta_k X_k) - G(\beta_0 + \beta_1 * 0 + \beta_2 X_2 + \ldots \beta_k X_k)$$

where **x_1 is binary explanatory variable** (for example gender) and other explanatory variables are fixed (at their means)

- We can use a following difference for other kind of discrete variables:

$$G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k (C_k + 1)) - G(\beta_0 + \beta_1 * 1 + \beta_2 X_2 + \ldots \beta_k C_k)$$

where **C_k is a number of children or credit score rating**

Some more on marginal effect

- It is straightforward to include standard functional forms among the explanatory variables. For example, in the model:

$$P(y = 1|z) = G(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_3 z_3)$$

- The partial effect **from change $\mathbf{z_1}$** on $P(y = 1|z)$ is:

  $\partial P(y = 1|z)/\partial z_1 = g(\beta_0 + x\beta)(\beta_1 + 2\beta_2 z_1)$, and

- The partial effect **from change $\mathbf{z_2}$** on $P(y = 1|z)$ is:

  $\partial P(y = 1|z)/\partial z_2 = g(\beta_0 + x\beta)(\beta_3/z_2)$

where $x\beta = \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_3 z_3$

Coefficients from LPV, probit and logit models

- Amemiya (1981) suggested the following relation between probit and logit:

$$\beta_{probit} \approx 0.625\,\beta_{logit}$$

- For the LPM and logit:

$$\beta_{LPV} \approx 0.25\,\beta_{logit} \text{ (except for the intercept)}$$
$$\beta_{LPV} \approx 0.25\,\beta_{logit} + 0.5 \text{ (for the intercept)}$$

Application1: Bank decision on the mortgage applications (LPM)

```
Linear regression                               Number of obs     =      2,380
                                                F(3, 2376)        =      59.63
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1216
                                                Root MSE          =     .30454
```

deny	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
pi_rat	.5269488	.0814237	6.47	0.000	.3672799	.6866177
black	.1331286	.0241857	5.50	0.000	.0857013	.1805559
ccred	.0427313	.0049558	8.62	0.000	.0330132	.0524494
_cons	-.1639722	.0267191	-6.14	0.000	-.2163674	-.111577

Application1: Bank decision on the mortgage applications (logit)

```
Iteration 0:   log pseudolikelihood =  -872.0853
Iteration 1:   log pseudolikelihood = -771.44022
Iteration 2:   log pseudolikelihood = -748.85275
Iteration 3:   log pseudolikelihood = -748.76599
Iteration 4:   log pseudolikelihood = -748.76597

Logistic regression                             Number of obs     =      2,380
                                                Wald chi2(3)      =     199.24
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -748.76597               Pseudo R2         =     0.1414
```

deny	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
pi_rat	5.133146	.9653946	5.32	0.000	3.241007	7.025285
black	.9499233	.1556954	6.10	0.000	.6447659	1.255081
ccred	.3402578	.0339427	10.02	0.000	.2737313	.4067842
_cons	-4.849078	.3575018	-13.56	0.000	-5.549769	-4.148387

Logit: Marginal Effects

	mean	$\hat{\beta}$	product
pi_rat	0.330814	5.133146	1.6981
black	0.142437	0.9499233	0.1353
ccred	2.116387	0 .3402578	0.7201
constant	1.00	-4.849078	–4.8491
Total			-2.2962

$$e^{-Z} = e^{-2.2962} = 0.100495$$

$$g(Z) = \frac{e^{-Z}}{(1+e^{-Z})^2}$$
$$= 0.082979$$

We then calculate $g(Z)$.

Logit: Marginal Effects

	mean	$\hat{\beta}$	product	$g(Z)$	$g(Z)\,\hat{\beta}$
pi_rat	0.330814	5.133146	1.6981	0.082979	0.4295
black	0.142437	0.9499233	0.1353	0.082979	0.0788
ccred	2.116387	0 .3402578	0.7201	0.082979	0.0282
constant	1.00	-4.849078	-4.8491		
Total			-2.2962		

$$\frac{\partial G}{\partial X_i} = \frac{\mathrm{d}G}{\mathrm{d}Z}\frac{\partial Z}{\partial X_i} = g(Z)\beta_i$$

The estimated marginal effects are $g(Z)$ multiplied by the respective coefficients.

Application1: Bank decision on the mortgage applications (probit)

```
Probit regression                               Number of obs     =      2,380
                                                Wald chi2(3)      =     204.11
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -749.37101               Pseudo R2         =     0.1407
```

deny	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
pi_rat	2.642005	.4382521	6.03	0.000	1.783046	3.500963
black	.5289367	.0870488	6.08	0.000	.3583241	.6995492
ccred	.1880319	.0188506	9.97	0.000	.1510854	.2249784
_cons	-2.65843	.1647258	-16.14	0.000	-2.981287	-2.335573

Probit: Marginal Effects

	mean	$\hat{\beta}$	product
pi_rat	0.330814	2.64201	0.8740
black	0.142437	0.52894	0.0753
ccred	2.116387	0.18803	0.3979
constant	1.00	-2.65843	–2.6584
Total			-1.3112

$$Z = \beta_1 + \beta_2 \bar{X}_2 + \ldots \beta_k \bar{X}_k = -1.3112$$

$$g(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} = 0.169147$$

In this case *Z* is equal to -1.3112 when the *X* variables are equal to their sample means.

We then calculate g(*Z*).

Probit: Marginal Effects

	mean	$\hat{\beta}$	product	$g(Z)$	$g(Z)\,\hat{\beta}$
pi_rat	0.330814	2.64201	0.8740	0.1691	0.4468
black	0.142437	0.52894	0.0753	0.1691	0.0894
ccred	2.116387	0.18803	0.3979	0.1691	0.0318
constant	1.00	-2.65843	–2.6584	0.1691	
Total			-1.3112		

$$\frac{\partial G}{\partial X_i} = \frac{\mathrm{d}G}{\mathrm{d}Z}\frac{\partial Z}{\partial X_i} = g(Z)\beta_i$$

The estimated marginal effects are $g(Z)$ multiplied by the respective coefficients.

	Logit	Probit
	$g(Z)\,\hat{\beta}$	$g(Z)\,\hat{\beta}$
pi_rat	0.4295	0.4468
black	0.0788	0.0894
ccred	0.0282	0.0318

The logit and probit results are displayed for comparison. The coefficients in the regressions are very different because different mathematical functions are being fitted. Nevertheless, the estimates of the marginal effects are usually similar.

However, if the outcomes in the sample are divided between a large majority and a small minority, they can differ. This is because the observations are then concentrated in a tail of the distribution. Although the logit and probit functions share the same sigmoid outline, their tails are somewhat different.

Application1: Marginal effects (probit, cont)

deny	dF/dx	Robust Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
pi_rat	.446225	.0725547	6.03	0.000	.330814	.30402	.58843
black*	.1127733	.02237	6.08	0.000	.142437	.068929	.156618
ccred	.0317579	.0032244	9.97	0.000	2.11639	.025438	.038078
obs. P	.1197479						
pred. P	.0949069	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Application1: Marginal effects (probit, cont.)

- *Prob(deny=1) = G(const, pi_rat, ccred, black)*

- Means:

variable	pi_rat	black	ccred
mean	0.330814	0.142437	2.116387

For binary: $G(\beta_0 + \boldsymbol{\beta_1} * \mathbf{1} + \beta_2 X_2 + \ldots \beta_k X_k)$ - $G(\beta_0 + \boldsymbol{\beta_1} * \mathbf{0} + \beta_2 X_2 + \ldots \beta_k X_k)$

- G (pi_rat = 0.330814, **black =0** , ccred=2.116387) =$\Phi(-2.66 + 2.64 * 0.33 + 0.53 * 0 + 0.19 * 2.12)$ =$\Phi(-1.386) = 1 - \Phi(1.386)$ = 0.0823

G (pi_rat = 0.330814, **black =1** , ccred=2.116387) =$\Phi(-2.66 + 2.64 * 0.33 + 0.53 * 1 + 0.19 * 2.12)$ =$\Phi(-0.856) = 1 - \Phi(0.856)$ = 0.1949

- **For binary var. BLACK** (marg. effect of black =0.1949-0.0823=0.1126)

Application1: Marginal effects (probit, cont.)

- *Prob(deny=1) = G(const, pi_rat, ccred, black)*

- Means:

variable	pi_rat	black	ccred
mean	0.330814	0.142437	2.116387

For discrete variable: $G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \boldsymbol{\beta_k}(\boldsymbol{C_k} + \boldsymbol{1}))$ - $G(\beta_0 + \beta_1 * 1 + \beta_2 X_2 + \ldots \boldsymbol{\beta_k C_k})$

- G (pi_rat = 0.330814, **black =0** , **ccred=2**) =$\Phi(-2.66 + 2.64 * 0.33 + 0.53 * 0 + 0.19 * 2)$ =$\Phi(-1.409) = 1 - \Phi(1.409)$ = 0.0681

G (pi_rat = 0.330814, **black =0**, **ccred=3**) =$\Phi(-2.66 + 2.64 * 0.33 + 0.53 * 0 + 0.19 * 3)$
=$\Phi(-1.219) = 1 - \Phi(1.219)$ = 0.1112

- **For discrete var. CCRED** (marg. effect of score declining from 2 to 3=0.1112-0.0881=0.0431)

TABLE 9.2 Mortgage Denial Regressions Using the Boston HMDA Data

Dependent Variable: *deny* = 1 If Mortgage Application Is Denied, = 0 If Accepted; 2,380 observations.						
Regression Model	***LPM***	***Logit***	***Probit***	***Probit***	***Probit***	***Probit***
Regressor	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
black	0.084** (0.023)	0.688** (0.182)	0.389** (0.098)	0.371** (0.099)	0.363** (0.100)	0.246 (0.448)
P/I ratio	0.449** (0.114)	4.76** (1.33)	2.44** (0.61)	2.46** (0.60)	2.62** (0.61)	2.57** (0.66)
housing expense-to-income ratio	−0.048 (.110)	−0.11 (1.29)	−0.18 (0.68)	−0.30 (0.68)	−0.50 (0.70)	−0.54 (0.74)
medium loan-to-value ratio (0.80 ≤ *loan-value ratio* ≤ 0.95)	0.031* (0.013)	0.46** (0.16)	0.21** (0.08)	0.22** (0.08)	0.22** (0.08)	0.22** (0.08)
high loan-to-value ratio (*loan-value ratio* ≥ 0.95)	0.189** (0.050)	1.49** (0.32)	0.79** (0.18)	0.79** (0.18)	0.84** (0.18)	0.79** (0.18)
consumer credit score	0.031** (0.005)	0.29** (0.04)	0.15** (0.02)	0.16** (0.02)	0.34** (0.11)	0.16** (0.02)
mortgage credit score	0.021 (0.011)	0.28* (0.14)	0.15* (0.07)	0.11 (0.08)	0.16 (0.10)	0.11 (0.08)
public bad credit record	0.197** (0.035)	1.23** (0.20)	0.70** (0.12)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)
denied mortgage insurance	0.702** (0.045)	4.55** (0.57)	2.56** (0.30)	2.59** (0.29)	2.59** (0.30)	2.59** (0.29)

(Table 9.2 continued)

TABLE 9.2 Mortgage Denial Regressions Using the Boston HMDA Data

Dependent Variable: *deny* = 1 If Mortgage Application Is Denied, = 0 If Accepted; 2,380 observations.

Regression Model **Regressor**	***LPM*** **(1)**	***Logit*** **(2)**	***Probit*** **(3)**	***Probit*** **(4)**	***Probit*** **(5)**	***Probit*** **(6)**
self-employed	0.060** (0.021)	0.67** (0.21)	0.36** (0.11)	0.35** (0.11)	0.34** (0.11)	0.35** (0.11)
single				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
high school diploma				−0.61** (0.23)	−0.60* (0.24)	−0.62** (0.23)
unemployment rate				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
condominium					−0.05 (0.09)	
black × *P/I ratio*						−0.58 (1.47)
black × *housing expense-to-income ratio*						1.23 (1.69)
Additional credit rating indicator variables	no	no	no	no	yes	no
constant	−0.183** (0.028)	−5.71** (0.48)	−3.04** (0.23)	−2.57** (0.34)	−2.90** (0.39)	−2.54** (0.35)

(Table 9.2 continued)

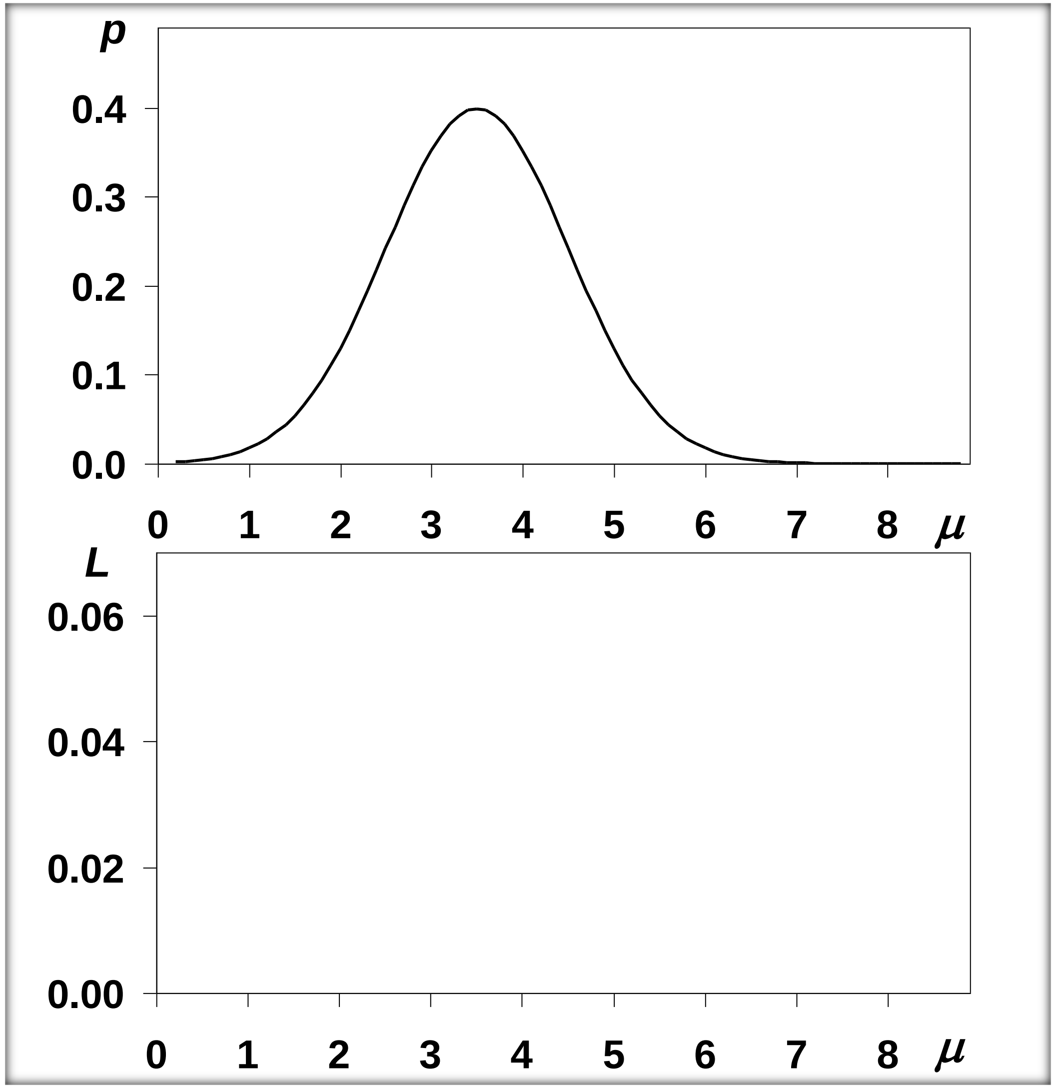
(Table 9.2 continued)

F-statistics and p-values Testing Exclusion of Groups of Variables

	(1)	(2)	(3)	(4)	(5)	(6)
Applicant single; HS diploma; industry unemployment rate				5.85 (<0.001)	5.22 (0.001)	5.79 (<0.001)
Additional credit rating indicator variables					1.22 (0.291)	
Race interactions and black						4.96 (0.002)
Race interactions only						0.27 (0.766)
Difference in predicted probability of denial, white vs. black (percentage points)	8.4%	6.0%	7.1%	6.6%	6.3%	6.5%

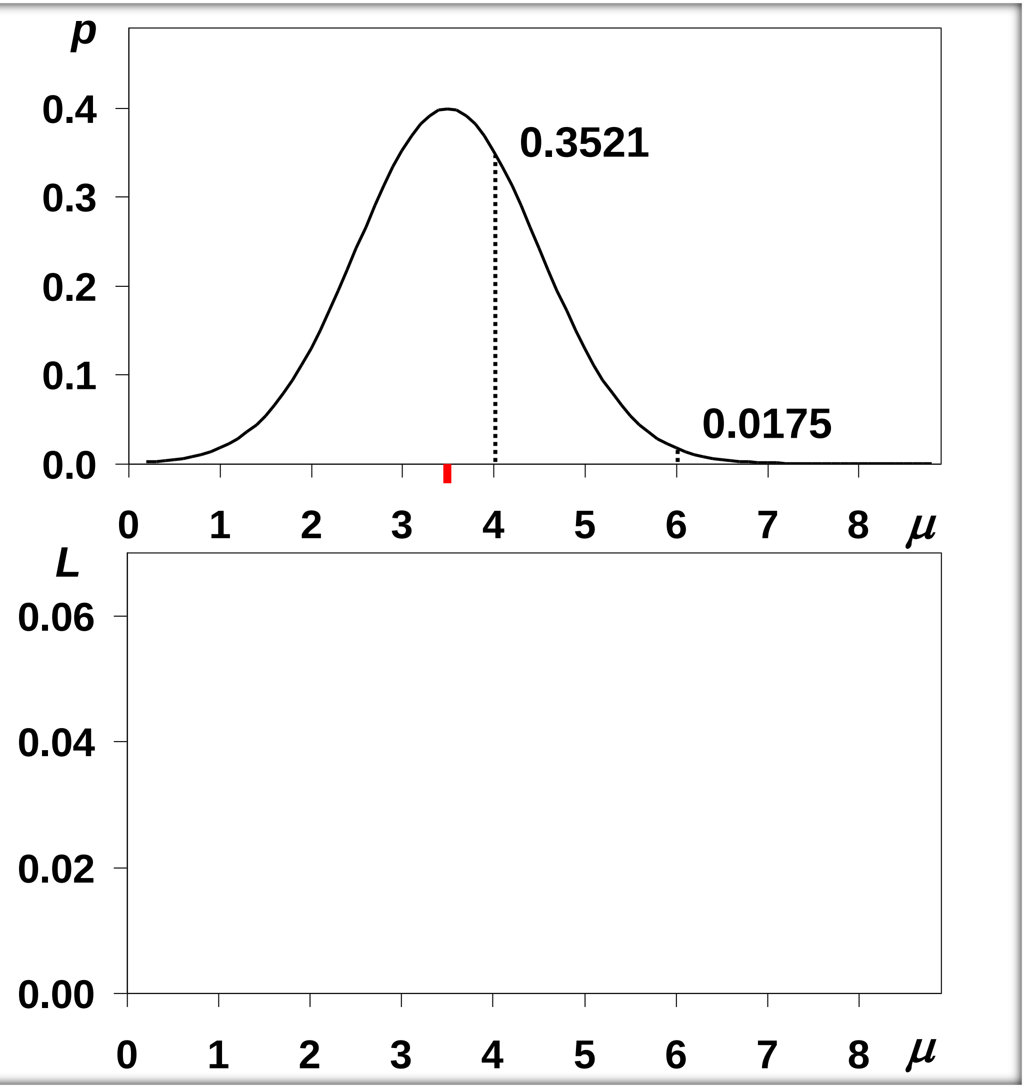
These regressions were estimated using the $n = 2{,}380$ observations in the Boston HMDA data set described in Appendix 9.1. The linear probability model was estimated by OLS, and probit and logit regressions were estimated by maximum likelihood. Standard errors are given in parentheses under the coefficients and p-values are given in parentheses under the F-statistics. The change in predicted probability in the final row was computed for a hypothetical applicant whose values of the regressors, other than race, equal the sample mean. Individual coefficients are statistically significant at the *5% or **1% level.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



This sequence introduces the principle of maximum likelihood estimation and illustrates it with some simple examples. Suppose that you have a normally-distributed random variable *X* with unknown population mean μ and standard deviation σ, and that you have a sample of two observations, 4 and 6. For the time being, we will assume that σ is equal to 1.

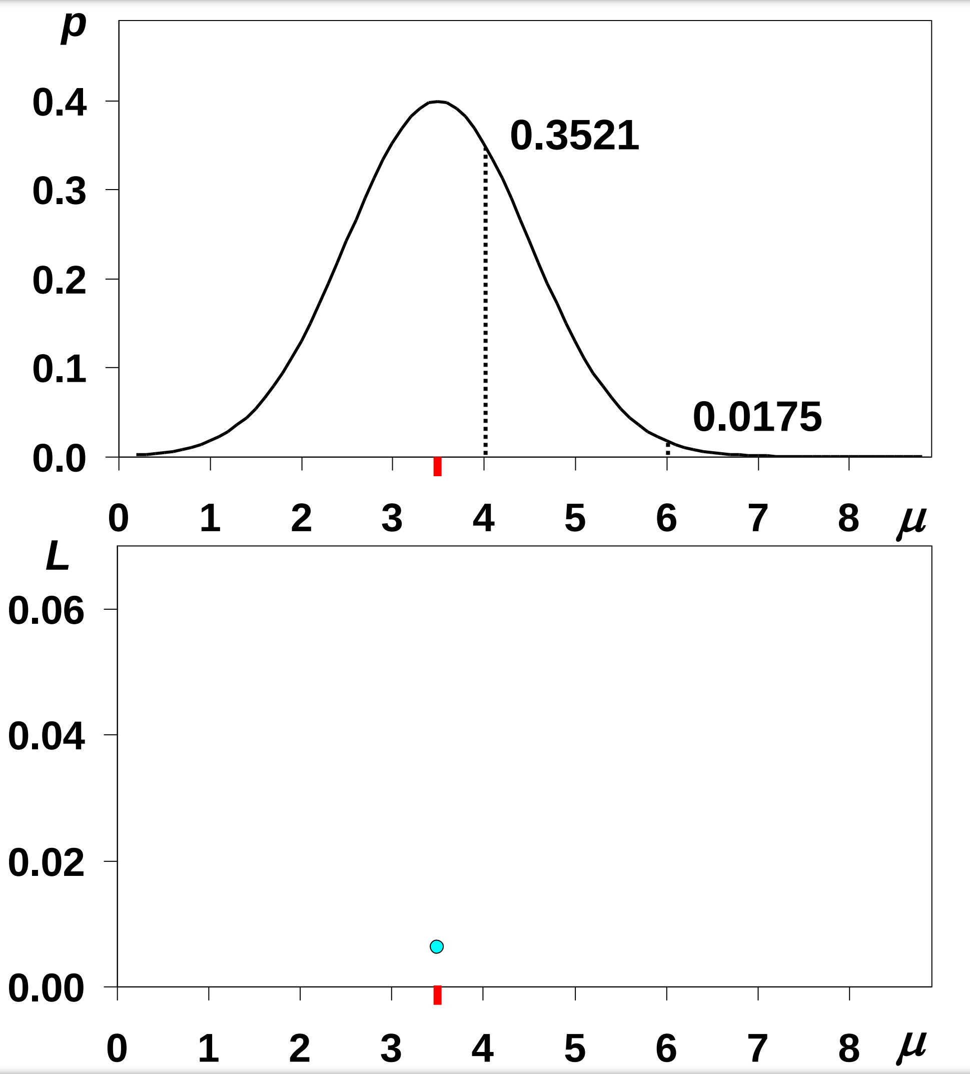
INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$
3.5	0.3521	0.0175

We want to obtain an estimate of μ. Suppose initially you consider the hypothesis $\mu = 3.5$. Under this hypothesis the probability density at 4 would be 0.3521 and that at 6 would be 0.0175.

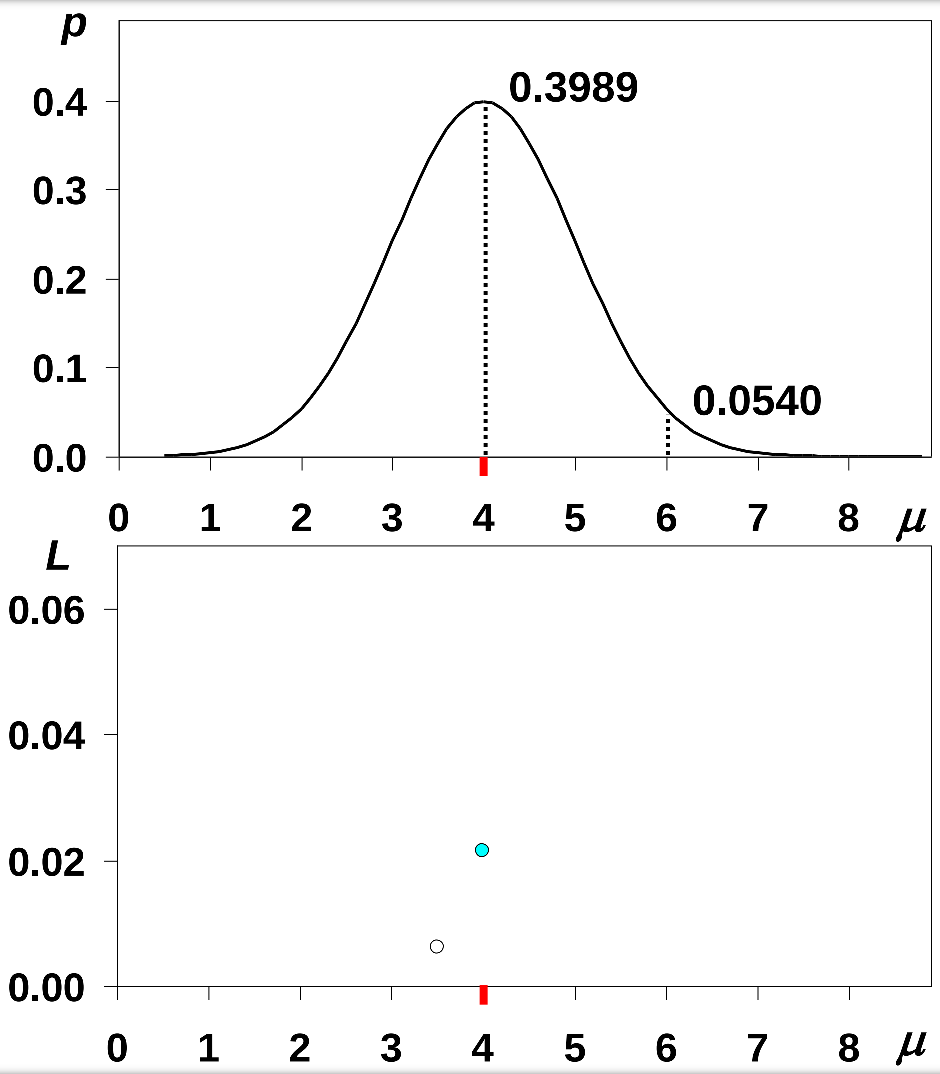
INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062

The joint probability density, shown in the bottom chart, is the product of these, 0.0062.

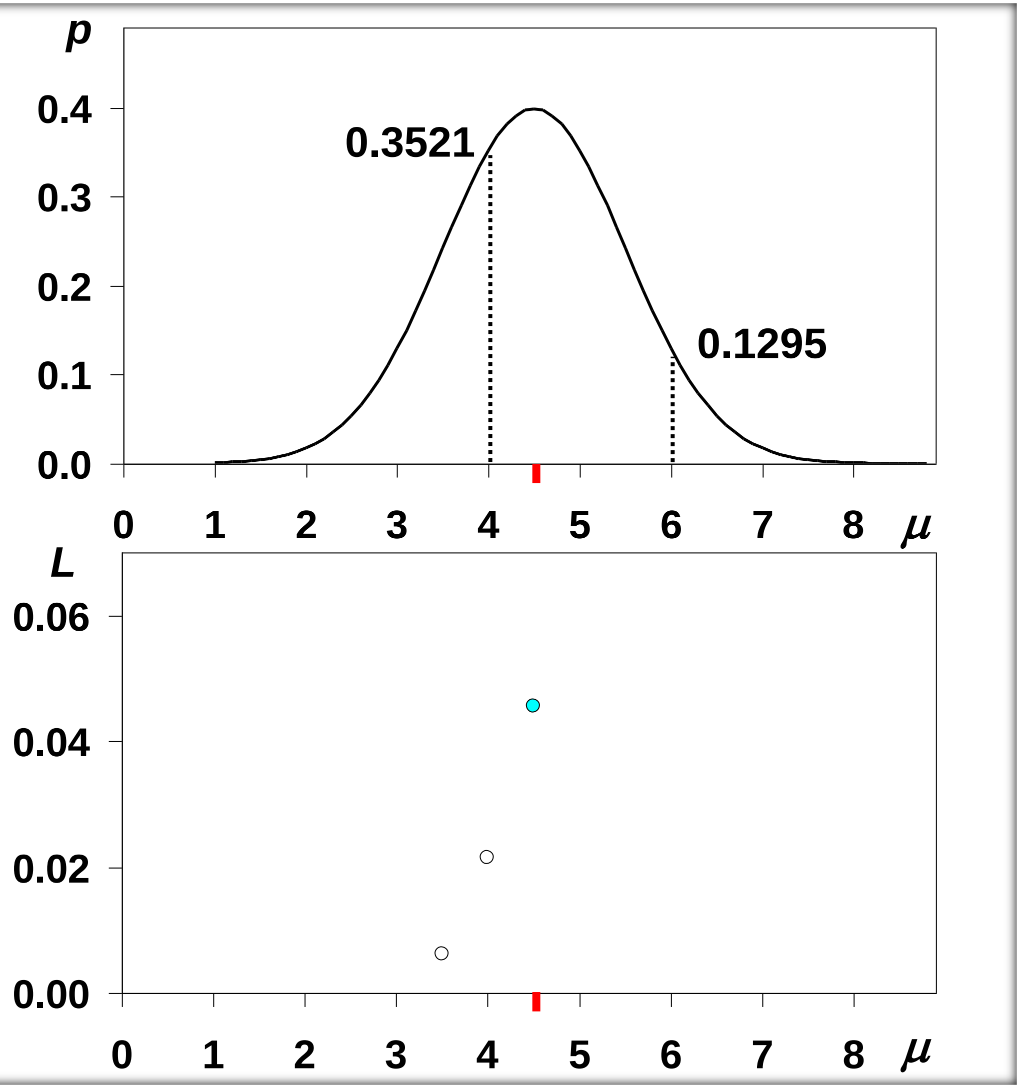
INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215

Next consider the hypothesis $\mu = 4.0$. Under this hypothesis the probability densities associated with the two observations are 0.3989 and 0.0540, and the joint probability density is 0.0215.

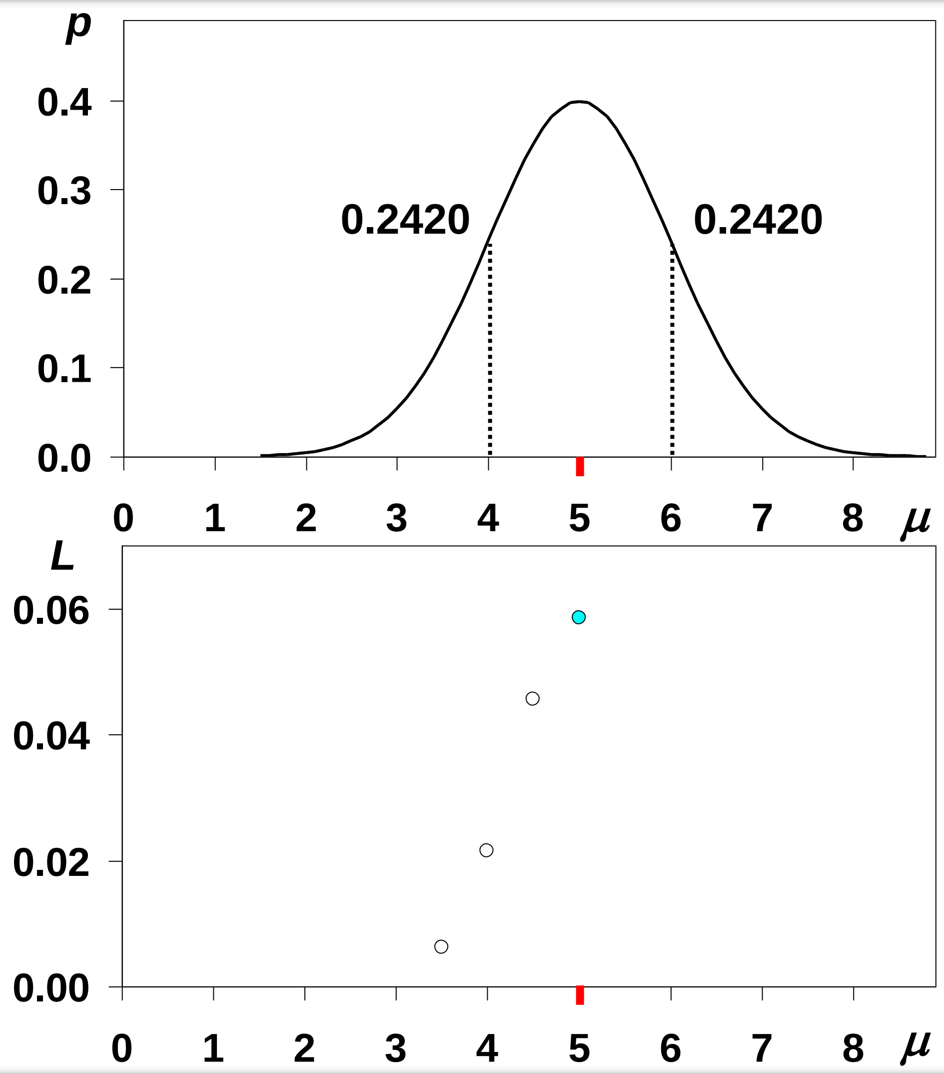
INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456

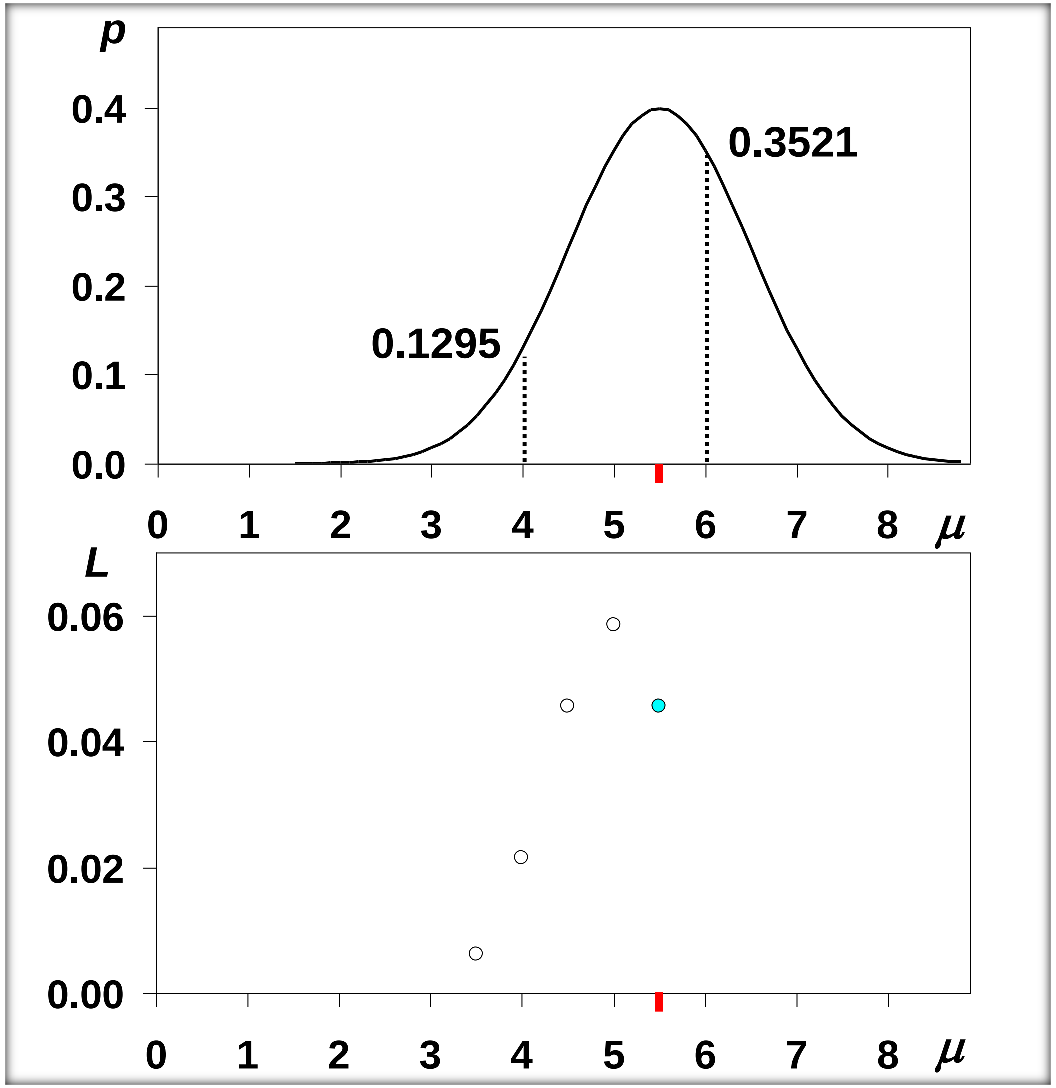
Under the hypothesis μ = 4.5, the probability densities are 0.3521 and 0.1295, and the joint probability density is 0.0456.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585

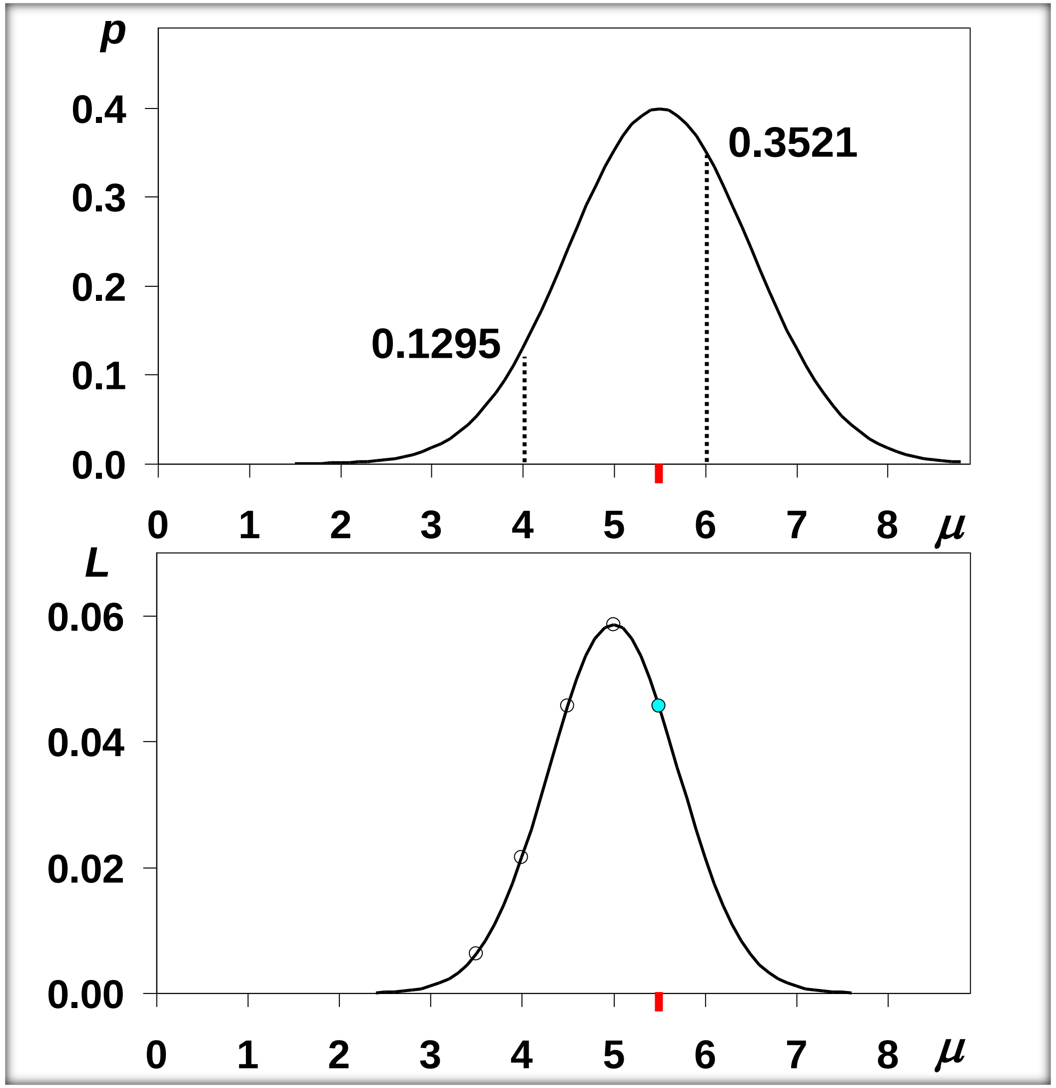
Under the hypothesis $\mu = 5.0$, the probability densities are both 0.2420 and the joint probability density is 0.0585.



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585
5.5	0.1295	0.3521	0.0456

Under the hypothesis $\mu = 5.5$, the probability densities are 0.1295 and 0.3521 and the joint probability density is 0.0456.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585
5.5	0.1295	0.3521	0.0456

The complete joint density function for all values of μ has now been plotted in the lower diagram. We see that it peaks at $\mu = 5$. MLE of μ is defined as value that maximizes the likelihood function given the sample data.

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

Now we will look at the mathematics of the example. If X is normally distributed with mean μ and standard deviation σ, its density function is as shown.

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2}$$

For the time being, we are assuming σ is equal to 1, so the density function simplifies to the second expression.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2}$$

$$f(4) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \qquad f(6) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}$$

Hence, we obtain the probability densities for the observations where *X* = 4 and *X* = 6.

$$f(X)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

$$f(X)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(X-\mu)^2}$$

$$f(4)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(4-\mu)^2} \qquad f(6)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(6-\mu)^2}$$

$$\text{joint density } =\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(4-\mu)^2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(6-\mu)^2}\right)$$

The joint probability density for the two observations in the sample is just the product of their individual densities.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

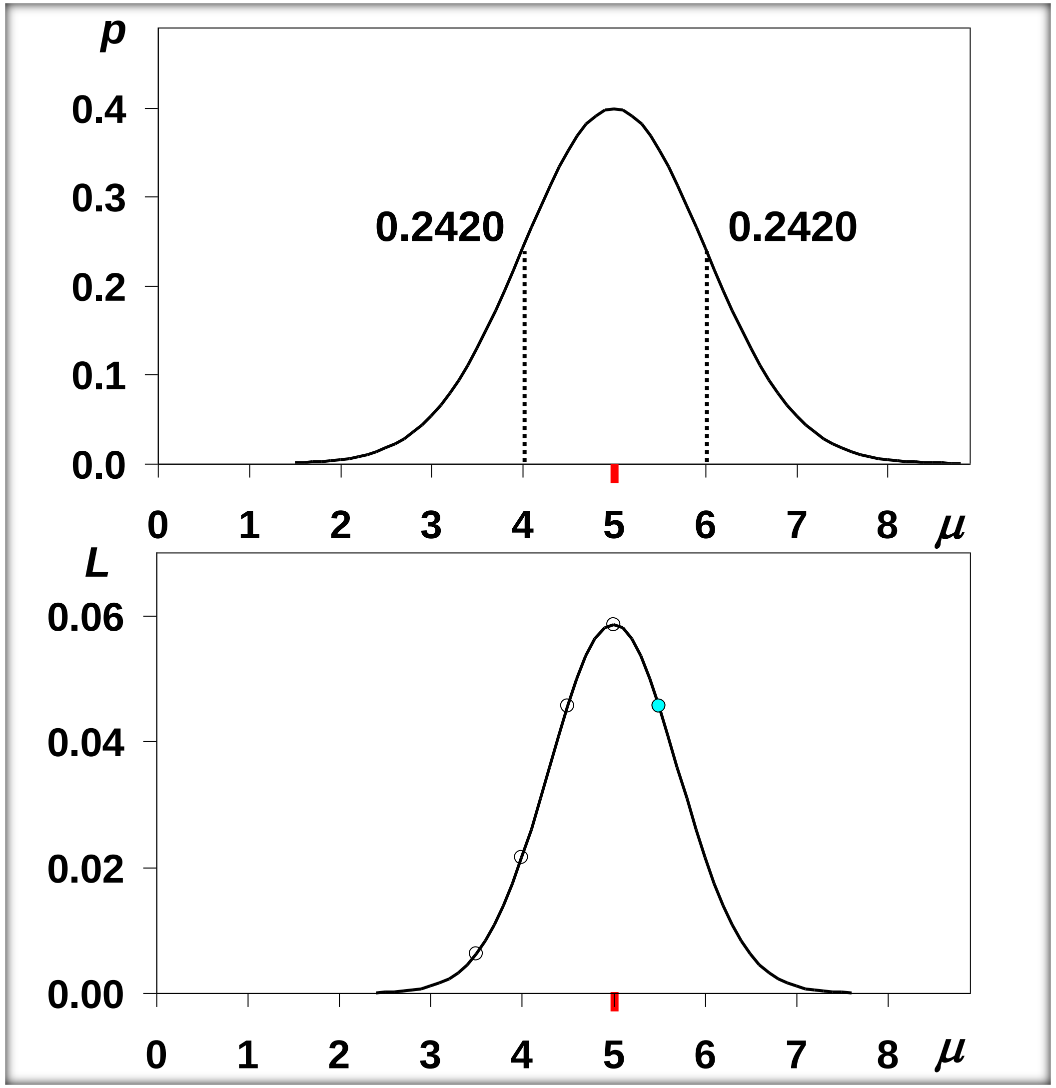
$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2}$$

$$f(4) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \qquad f(6) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}$$

$$\text{joint density} = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}\right)$$

In maximum likelihood estimation we choose as our estimate of μ the value that gives us the greatest joint density for the observations in our sample. This value is associated with the greatest probability, or maximum likelihood, of obtaining the observations in the sample.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION



μ	$p(4)$	$p(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585
5.5	0.1295	0.3521	0.0456

In the graphical treatment we saw that this occurs when μ is equal to 5. We will prove this must be the case mathematically.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

To do this, we treat the sample values *X* = 4 and *X* = 6 as given and we use the calculus to determine the value of μ that maximizes the expression.

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

When it is regarded in this way, the expression is called the likelihood function for μ given the sample observations 4 and 6. This is the meaning of $L(\mu \mid 4,6)$.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

To maximize the expression, we could differentiate with respect to μ and set the result equal to 0. This would be a little laborious. Fortunately, we can simplify the problem with a trick.

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

log *L* is a monotonically increasing function of *L* (meaning that log *L* increases if *L* increases and decreases if *L* decreases).

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

It follows that the value of μ which maximizes log *L* is the same as the one that maximizes *L*. As it so happens, it is easier to maximize log *L* with respect to μ than it is to maximize *L*.

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(6-\mu)^2} \right)$$

The logarithm of the product of the density functions can be decomposed as the sum of their logarithms. Using the product rule a second time, we can decompose each term as shown.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log\left(\frac{1}{\sqrt{2\pi}} \right) + \log\left(e^{-\frac{1}{2}(4-\mu)^2} \right) + \log\left(\frac{1}{\sqrt{2\pi}} \right) + \log\left(e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$= 2\log\left(\frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

$$\log a^b = b \log a$$

$$\log e^{-\frac{1}{2}(X-4)^2} = -\frac{1}{2}(X-4)^2 \log e = -\frac{1}{2}(X-4)^2$$

Hence the second term reduces to a simple quadratic in *X*. And so does the fourth.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log\left(\frac{1}{\sqrt{2\pi}} \right) + \log\left(e^{-\frac{1}{2}(4-\mu)^2} \right) + \log\left(\frac{1}{\sqrt{2\pi}} \right) + \log\left(e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$= 2\log\left(\frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

We will now choose μ so as to maximize this expression.

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

$$-\frac{1}{2}(a-\mu)^2 = -\frac{1}{2}\left(a^2 - 2a\mu + \mu^2\right) = -\frac{1}{2}a^2 + a\mu - \frac{1}{2}\mu^2$$

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

$$-\frac{1}{2}(a-\mu)^2 = -\frac{1}{2}\left(a^2 - 2a\mu + \mu^2\right) = -\frac{1}{2}a^2 + a\mu - \frac{1}{2}\mu^2$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left\{-\frac{1}{2}(a-\mu)^2\right\} = a - \mu$$

Quadratic terms of the type in the expression can be expanded as shown.

Thus, we obtain the differential of the quadratic term.

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

$$-\frac{1}{2}(a-\mu)^2 = -\frac{1}{2}\left(a^2 - 2a\mu + \mu^2\right) = -\frac{1}{2}a^2 + a\mu - \frac{1}{2}\mu^2$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left\{-\frac{1}{2}(a-\mu)^2\right\} = a - \mu$$

$$\frac{\mathrm{d}\log L}{\mathrm{d}\mu} = (4-\mu) + (6-\mu)$$

Applying this result, we obtain the differential of log L with respect to μ. (The first term in the expression for log L disappears completely since it is not a function of μ.)

INTRODUCTION TO MAXIMUM LIKELIHOOD ESTIMATION

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

$$-\frac{1}{2}(a-\mu)^2 = -\frac{1}{2}\left(a^2 - 2a\mu + \mu^2\right) = -\frac{1}{2}a^2 + a\mu - \frac{1}{2}\mu^2$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left\{-\frac{1}{2}(a-\mu)^2\right\} = a - \mu$$

$$\frac{\mathrm{d}\log L}{\mathrm{d}\mu} = (4-\mu) + (6-\mu)$$

$$\frac{\mathrm{d}\log L}{\mathrm{d}\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = 5$$

Thus, from the first order condition we confirm that 5 is the value of μ that maximizes the log-likelihood function, and hence the likelihood function.

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

$$-\frac{1}{2}(a-\mu)^2 = -\frac{1}{2}\left(a^2 - 2a\mu + \mu^2\right) = -\frac{1}{2}a^2 + a\mu - \frac{1}{2}\mu^2$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left\{-\frac{1}{2}(a-\mu)^2\right\} = a - \mu$$

$$\frac{\mathrm{d}\log L}{\mathrm{d}\mu} = (4-\mu) + (6-\mu)$$

$$\frac{\mathrm{d}\log L}{\mathrm{d}\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = 5$$

$$\frac{\mathrm{d}^2\log L}{\mathrm{d}\mu^2} = -2$$

Note also that the second differential of log *L* with respect to μ is –2. Since this is negative, we have found a maximum, not a minimum.

Maximum Likelihood Estimation of Logit and Probit Models

- MLE is **based on the distribution on given x**, the heteroscedasticity in $Var(y|x)$ is automatically accounted for

- To obtain the maximum likelihood estimator, conditional on the explanatory variables, we need the density of y_i given x_i (*n*-sample size). We can write this as:

$$f(y_i|x_i;\beta)= [G(x_i\beta)]^{y_i}[1 - G(x_i\beta)]^{1-y_i}, \quad y=0,1$$

- The log-likelihood function for observation *i* can be obtained:

$$l_i(\beta)=y_i log[G(x_i\beta)]+(1- y_i)log[1 - G(x_i\beta)]$$

- Because $G[.]$ is strictly between zero and one for logit and probit, $l_i(\beta)$ is well defined for all values of β

Maximum Likelihood Estimation of Logit and Probit Models, cont.

- The log-likelihood for a sample size n is obtained by summing log-likelihood of all *n* observations:

$$\log L_n = \log L_n\,(x;\beta) = \mathcal{L}(\beta) = \sum_{i=1}^{n} l_i(\beta) = \sum_{i=1}^{n} \{y_i \log[\mathrm{G}(x_i\beta] + (1-y_i)\log[1-\mathrm{G}(x_i\beta)]\}$$

- The MLE of β, denoted as $\hat{\beta}$ maximizes this log-likelihood (it is easy to show that: $\partial L(x;\beta) / \partial\beta = 0$, $\partial^2 L(x;\beta) / \partial\beta\, \partial\beta' < 0$)

- The general theory of MLE for random samples implies that, under very general conditions, the MLE **is consistent, asymptotically normal, and asymptotically efficient**. We will just use the results here

- Each $\hat{\beta}_j$ comes with an (asymptotic) standard error, the formula for which is complicated and presented in the chapter appendix. We form the t (z) statistic and carry out the test in the usual way, once we have decided on a one- or two-sided alternative

The Likelihood Ratio Test (Testing Multiple Hypothesis)

- Unlike the LPM, where we can compute *F* statistics or *LM* statistics to test exclusion restrictions, we need a new type of test
- Maximum likelihood estimation (MLE), will always produce a log-likelihood, L
- Just as in an F test, you estimate the restricted and unrestricted model, then form
- $LR = 2(L_{ur} – L_r) \sim \chi^2_q$ (q number of restrictions)

Application1: Bank decision on the mortgage applications (probit, unrestricted)

```
Probit regression                                   Number of obs     =      2,380
                                                    Wald chi2(3)      =     204.11
                                                    Prob > chi2       =     0.0000
Log pseudolikelihood = -749.37101                   Pseudo R2         =     0.1407
```

deny	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
pi_rat	2.642005	.4382521	6.03	0.000	1.783046	3.500963
black	.5289367	.0870488	6.08	0.000	.3583241	.6995492
ccred	.1880319	.0188506	9.97	0.000	.1510854	.2249784
_cons	-2.65843	.1647258	-16.14	0.000	-2.981287	-2.335573

Application1: Bank decision on the mortgage applications (probit, restricted)

```
Probit regression                               Number of obs     =      2,380
                                                Wald chi2(0)      =          .
                                                Prob > chi2       =          .
Log pseudolikelihood =  -872.0853               Pseudo R2         =     0.0000
```

deny	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	-1.176248	.0333247	-35.30	0.000	-1.241563	-1.110933

.

LR statistic	245.4285
Prob(LR statistic)	0.0000

Goodness of Fit

- Unlike the LPM, where we can compute an R^2 to judge goodness of fit, we need **new measures of goodness of fit**

- One possibility is a **pseudo R^2** based on the log likelihood and defined as $1 – L_{ur}/L_r$

- Can also look at the percent correctly predicted – if predict a probability >0.5 then that matches y = 1 and vice versa

Application1: Bank decision on the mortgage applications - Percent correctly predicted from probit

Classified	True D	True ~D	Total
+	25	26	51
-	260	2069	2329
Total	285	2095	2380

Classified + if predicted Pr(D) >= .5
True D defined as deny != 0

Sensitivity	Pr(+| D)	8.77%
Specificity	Pr(-|~D)	98.76%
Positive predictive value	Pr(D| +)	49.02%
Negative predictive value	Pr(~D| -)	88.84%
False + rate for true ~D	Pr(+|~D)	1.24%
False - rate for true D	Pr(-| D)	91.23%
False + rate for classified +	Pr(~D| +)	50.98%
False - rate for classified -	Pr(D| -)	11.16%
Correctly classified		87.98%

Application1: Bank decision on the mortgage applications – Behind the percent correctly predicted from probit

no. obs.	deny	predictions	true of not
1	1	0.8299	yes
2	0	0.8189	no
3	1	0.7239	yes
4	0	0.7210	no
5	1	0.5744	yes
6	1	0.5640	yes
7	0	0.5616	no
8	1	0.5536	yes
9	1	0.5438	yes
10	0	0.4431	yes
11	0	0.4392	yes
12	1	0.5326	yes
13	0	0.5316	no
14	1	0.4905	no
15	0	0.4893	yes

Application 2: MROZ data (married women LFP, Y=inlf – in labor force or not)

TABLE 17.1 LPM, Logit, and Probit Estimates of Labor Force Participation

Dependent Variable: *inlf*			
Independent Variables	LPM (OLS)	Logit (MLE)	Probit (MLE)
nwifeinc	−.0034 (.0015)	−.021 (.008)	−.012 (.005)
educ	.038 (.007)	.221 (.043)	.131 (.025)
exper	.039 (.006)	.206 (.032)	.123 (.019)
$exper^2$	−.00060 (.00019)	−.0032 (.0010)	−.0019 (.0006)
age	−.016 (.002)	−.088 (.015)	−.053 (.008)
kidslt6	−.262 (.032)	−1.443 (.204)	−.868 (.119)
kidsge6	.013 (.014)	.060 (.075)	.036 (.043)
constant	.586 (.152)	.425 (.860)	.270 (.509)
Percentage correctly predicted	73.4	73.6	73.4
Log-likelihood value	—	−401.77	−401.30
Pseudo R-squared	.264	.220	.221

Application 2: MROZ data (married women LFP)

TABLE 17.2 Average Partial Effects for the Labor Force Participation Models

Independent Variables	LPM	Logit	Probit
nwifeinc	–.0034 (.0015)	–.0038 (.0015)	–.0036 (.0014)
educ	.038 (.007)	.039 (.007)	.039 (.007)
exper	.027 (.002)	.025 (.002)	.026 (.002)
age	–.016 (.002)	–.016 (.002)	–.016 (.002)
kidslt6	–.262 (.032)	–.258 (.032)	–.261 (.032)
kidsge6	.013 (.014)	.011 (.013)	.011 (.013)

Application 2: MROZ data (married women LFP)

FIGURE 17.2 Estimated response probabilities with respect to education for the linear probability and probit models.