University College London
Department of Economics

G023: Econometric Theory and Methods
Exercise 1: Sketch answers*

1. This problem characterizes the expected value as the best predictor in the sense of minimizing the mean squared error. Note that the values of a and b are irrelevant so long as $b > 0$ because

$$E_X[a + b(X - x^*)^2] = a + bE_X\left[(X - x^*)^2\right].$$

A direct attack gives on differentiating the objective function

$$\frac{\partial}{\partial x^*}E_X[(X - x^*)^2] = -2bE_X[(X - x^*)]$$

which is zero uniquely at $x^* = E_X[X]$ when $b > 0$. Differentiating again gives a value of $2b$ which is positive indicating that we have our loss minimising predictor. Part (a) of the question leads to

$$E_X[a + b(X - x^*)^2] = a + bVar_X(X) + b\left(E_X[X] - x^*\right)^2$$

which is clearly minimised by setting $x^* = E_X[X]$.

The observation generalizes to the case in which X has a distribution conditional on the value of other variables, Z. The best conditional predictor of X in the sense of minimizing the conditional mean squared error (conditional on $Z = z$, say) is the conditional mean function $E\left(X|Z = z\right)$.

If we change the criterion from the mean squared error to something else, then the mean may no longer be the best predictor. For example, if we use the absolute deviation criterion, $a + bE_X\left[|X - x^*|\right]$then the best predictor is the median of X. Of course for some distributions (e.g. symmetric distributions) the mean and median may be equal.

*Hidehiko Ichimura contributed to these answers.

2. This problem reviews density and distribution functions, and moment generating functions and their conditional counter-parts. A scedastic function describes how the variance of a conditional distribution changes with the value of the conditioning variable. You need to be able to compute some simple integrals in this question.

   (a) Differentiating $F_Y(y)$ with respect to y gives $f_Y(y) = \lambda \exp(-\lambda y)$.

   (b) $M_Y(t) = E_Y[\exp(tY)] = \int_0^\infty \lambda \exp(-(\lambda - t)y)dy = \lambda/(\lambda - t)$. Note, we require $\lambda > t$ which is fine because to get moments we will only examine the behaviour of $M_Y(t)$ in a neighbourhood of $t = 0$ (In this problem the tail of the density can go down fast enough (when $\lambda > t$) so that the moment generating function exists. In general the moment generating function may not exist but the characteristic function always exists. In this sense better to use characteristic functions if we want to use a method that always works. But that requires understanding some elements of complex analysis. Differentiating with respect to t and setting $t = 0$, gives $E_Y[Y] = 1/\lambda$, differentiating a second time and setting $t = 0$ gives $E_Y[Y^2] = 2/\lambda^2$ and using $Var[Y] = E_Y[Y^2] - E_Y[Y]^2$ gives $Var(Y) = 1/\lambda^2 = E_Y[Y]^2$.

   (c) Median unemployment duration, $Q_Y(\frac{1}{2})$, is the solution to

$$\frac{1}{2} = 1 - \exp(-\lambda Q_Y(\frac{1}{2}))$$

   that is

$$Q_Y(\frac{1}{2}) = -\frac{1}{\lambda}\log(\frac{1}{2})$$

   so when $\lambda = 2$, $Q_Y(\frac{1}{2}) = 0.347$ which is 127 days (recall Y is measured in years). The lower and upper quartiles are

$$\begin{aligned} Q_Y(\frac{1}{4}) &= -\frac{1}{\lambda}\log(\frac{3}{4}) \\ Q_Y(\frac{3}{4}) &= -\frac{1}{\lambda}\log(\frac{1}{4}) \end{aligned}$$

   and with $\lambda = 2$ we have

$$\begin{aligned} Q_Y(\frac{1}{4}) &= 0.144 \\ Q_Y(\frac{3}{4}) &= 0.693 \end{aligned}$$

   and so the interquartile range is

$$Q_Y(\frac{3}{4}) - Q_Y(\frac{1}{4}) = 0.549.$$

   For larger (smaller) values of λ the median unemployment duration and the interquartile range are smaller (larger). The expected value of unemployment duration is λ^{-1} which is 0.5 when $\lambda = 2$, larger than the median unemployment duration. Both sketch graphs should show hyperbolas, that for the median lying higher than that for the mean.

(d) The distribution function of $S = \log(Y)$ is

$$F_S(s) = P[S \leq s] = P[\log(Y) \leq s] = P[Y \leq \exp(s)] = F_Y(\exp(s)) = 1-\exp(-\lambda \exp(s)).$$

The probability density function of S is, on differentiating $F_S(s)$ with respect to s,

$$f_S(s) = \lambda \exp(s - \lambda \exp(s)).$$

Median log unemployment duration, $Q_S(\frac{1}{2})$ satisfies

$$\frac{1}{2} = 1 - \exp(-\lambda \exp(Q_S(\frac{1}{2})))$$

that is

$$Q_S(\frac{1}{2}) = \log(-\frac{1}{\lambda} \log(\frac{1}{2}))$$

and we already have

$$Q_Y(\frac{1}{2}) = -\frac{1}{\lambda} \log(\frac{1}{2})$$

and it is clear that

$$Q_S(\frac{1}{2}) = \log(Q_Y(\frac{1}{2})).$$

(e) By reinterpreting the meaning of λ as the conditional mean and substituting $\lambda = \alpha \exp(\beta x)$ in place of λ, we have $E_Y[Y|X = x] = \alpha^{-1} \exp(-\beta x)$, $Var[Y|X = x] = \alpha^{-2} \exp(-2\beta x)$. A scatter of points around an exponential function falling as x becomes large with more dispersion at small values of x (years of schooling) than at large values of x because α and $\beta > 0$. Of course all realisations would be non-negative.

3. This problem is an exercise on deriving a distribution of a random variable from the joint distribution of underlying random variables. The specific problem here is a special case of the more general problem of calculating a joint distribution of order statistics. Order statistics appear also in studying auction models and also in estimation of an edge of the support of a random variable.

This problem also examines your understanding about the quantile function. This is a useful concept that is increasingly used in the empirical literature.

(a)

$$\begin{aligned} F_Z[z] &= P[Z \leq z] \\ &= P[(Y_1 \leq z) \cap (Y_2 \leq z) \cap \cdots \cap (Y_n \leq z)] \\ &= P[Y \leq z]^n \\ &= F_Y[z]^n. \end{aligned}$$

At the third line we have used the independence property.

(b) Differentiating with respect to z gives

$$f_Z(z) = n f_Y(z) F_Y(z)^{n-1}.$$

(c) The p quantile of Z is the value z that satisfies $F_Z[z] = p$. Since $F_Z[z] = F_Y[z]^n$, the same solution that satisfies $F_Z[z] = p$ satisfies $F_Y[z]^n = p$ and thus $F_Y[z] = p^{1/n}$. But the z that satisfies the last expression is the $p^{1/n}$ quantile of Y. We have shown that the p quantile of Z is the same as the $p^{1/n}$ quantile of Y.

4. When we model choice variables, we can code them how we like.. We often use the values 0 and 1 because they are convenient for reasons you will see in this problem. If we denote the choices by $Y = 0$ and 1, then as we will see, $E[Y|X = x] = p(x)$ so that by defining $U = Y - p(X)$, we have $Y = p(X) + U$ where $E[U|X = x] = 0$ and $Var[U|X = x] = p(x)(1 - p(x))$.

(a)

$$E[Y|X = x] = 1 \times p(x) + 0 \times (1 - p(x)) = p(x).$$

(b)

$$E[Y^2|X = x] = 1^2 \times p(x) + 0^2 \times (1 - p(x)) = p(x),$$
$$Var[Y|X = x] = E[Y^2|X = x] - E[Y|X = x]^2 = p(x)(1 - p(x)).$$

(c) Here X needs to be interpreted as a vector of random variables X_1, X_2, ..., X_M and $X = x$ means that all of these random variables are fixed at the same particular value x. Let the random variable for the mth household be Y_m, $m = 1, \ldots, M$. Then $Z = (Y_1 + Y_2 + \cdots + Y_M)/M$.

i.

$$\begin{aligned}
E[Z|X_1 = x, ..., X_M = x] &= E[(Y_1 + Y_2 + \cdots + Y_M)/M|X_1 = x, ..., X_M = x] \\
&= \frac{1}{M}\{E[Y_1|X_1 = x] + \cdots + E[Y_M|X_M = x]\} \\
&= \frac{1}{M} M p(x) \\
&= p(x).
\end{aligned}$$

At the second line the restriction $E[Y_i|X_i = x] = E[Y_i|X_1 = x, ..., X_M = x]$ has been imposed.[1] To compute the variance,

$$\begin{aligned}
Var[Z|X_1 &= x, ..., X_M = x] \\
&= E[(Z - E[Z|X_1 = x, ..., X_M = x])^2|X_1 = x, ..., X_M = x] \\
&= \frac{1}{M^2} E\left\{[(Y_1 - p(x)) + \cdots + (Y_M - p(x))]^2 |X_1 = x, ..., X_M = x\right\}
\end{aligned}$$

[1] That restriction should, strictly speaking, have been made explicit in the question - it does not follow directly from the stated requirement that households' choices be independent. Independence of households' choices is a feature of the distribution of the Y's conditional on X's whereas this restriction is a restriction on the dependence of each Y_i on values of X's.

Expand the quadratic and note that when $i \neq j$ exploiting independence

$$\begin{aligned}
& E\left[\left(Y_i - p(x)\right)\left(Y_j - p(x)\right)|X_1 = x, ..., X_M = x\right] \\
= \; & E\left[\left(Y_i - p(x)\right)|X_1 = x, ..., X_M = x\right] E\left[\left(Y_j - p(x)\right)|X_1 = x, ..., X_M = x\right] \\
= \; & E\left[\left(Y_i - p(x)\right)|X_i = x\right] E\left[\left(Y_j - p(x)\right)|X_j = x\right] = 0
\end{aligned}$$

and

$$\begin{aligned}
& E\left[\left(Y_i - p(x)\right)^2 |X_1 = x, ..., X_M = x\right] \\
= \; & E\left[\left(Y_i - p(x)\right)^2 |X_i = x\right] = p\left(x\right)\left[1 - p\left(x\right)\right]
\end{aligned}$$

leading to

$$\begin{aligned}
Var[Z|X_1 = x, ..., X_M = x] &= \frac{1}{M^2} Mp(x)(1 - p(x)) \\
&= \frac{1}{M} p(x)(1 - p(x)).
\end{aligned}$$

ii. The variance is small when $p\left(x\right)$ is close to 0 or when it is close to 1 so under the assumption that $p\left(x\right)$ is increasing in income, large and small values of income both lead to a small conditional variance of Z.

iii. Dispersed scatter at medium levels of income, tight scatter at high and low levels of income, all around some kind of increasing line rising from zero to one. All points between zero and one.

5. *Set identification.*

(a) There is the Law of Total Probability:

$$E[Y_1] = E[Y_1|D = 1]P[D = 1] + E[Y_1|D = 0]P[D = 0] \qquad (1)$$

and on the right hand side data are informative about $E[Y_1|D = 1]$, $P[D = 1]$ and $P[D = 0]$, but not about $E[Y_1|D = 0]$. One could estimate the two probabilities using sample proportions of participants and non-participants. One could estimate $E[Y_1|D = 1]$ using the sample average of realised values of Y_1 for participants. We don't see realised values of Y_1 for non-participants so there is no way of estimating $E[Y_1|D = 0]$. There is a similar argument for $E[Y_0]$.[2]

(b) See Question 4 part (a) and apply that argument conditioning throughout on $D = 0$.

(c) In the equation (1) $E[Y_1]$ is an increasing function of the unknown $E[Y_1|D = 0]$ and so is at least equal to the value obtained when $E[Y_1|D = 0]$ is replaced by its smallest possible value (0) and can be no larger than the value obtained when it is replaced by its largest possible value (1). This gives the required result.

[2]This is a loose argument appealing to estimability rather than directly to identification. A formal argument will note that we see realisations of $Y = DY_1 + (1 - D)Y_0$ and that from knowledge of the joint distribution of Y and D one cannnot obtain knowledge of e.g. $E[Y_1]$.

(d) Suppose that $P(D = 1)$ is neither zero nor one. Recall that $E[Y_1]$ is a probability and so it must lie in $[0, 1]$. $E[Y_1|D = 1]$ is also a probability and lies in $[0, 1]$. The left hand side of the inequality is therefore a product of probabilities and as long as neither is zero the left hand side is greater than zero. Using the shorthand notation $a = E[Y_1|D = 1]$ and $p = P[D = 1]$, the right hand side of the inequality can be written

$$r = ap + 1 - p = 1 - (1 - a)p$$

and if $a < 1$ this is smaller than 1. The inequality is informative except in extreme cases because it defines a proper subset of $[0, 1]$.

(e) Now $0 \leq E[Y_1|D = 0] \leq E[Y_1|D = 1]$ and so interval of set identification becomes

$$E[Y_1|D = 1]P[D = 1] \leq E[Y_1] \leq E[Y_1|D = 1]P[D = 1]+E[Y_1|D = 1]P[D = 0]$$

which is clearly shorter, and simplifies to:

$$E[Y_1|D = 1]P[D = 1] \leq E[Y_1] \leq E[Y_1|D = 1].$$