Introduction to ECON4135

Review of elementary probability

Tore Schweder

August 2009

Goal

- Statistics and econometrics is methodology for **learning** from **data**
- Data are **systematically** obtained observations or measurements
- The **phenomenon** and/or the **data generating mechanism** might be **stochastic**
- Statistical models are cast in **probability** terms to capture **random**/stochastic variation (brief review of probability now)
- The goal of the course (and the book) is to make econometrics come alive

Random variables

Denoted by capital letters X, Y, Z... represent the outcome of experiments/observations, are discrete or continuously distributed on nominal/ordinal/interval scales.

Examples

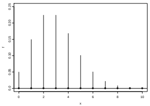
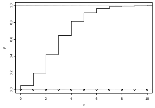
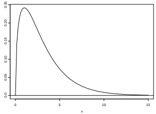
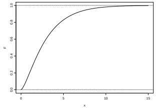
	Nominal	Ordinal	Interval
Discrete	gender	utility	utility number of children
Continuous	color	utility	utility value in NOK rate of inflation

Probability, density, cdf

X has a probability distribution characterized by probability function f or its cumulated version F

	f	F
Discrete	$P(X=x)=f(x)$	$F(x)=P(X\leq x)=\sum_{i\leq x} f(i)$
Continuous	$f(x)=\frac{d}{dx}F(x)$	$F(x)=P(X\leq x)=\int_{-\infty}^{x} f(u)\,du$

	Expected value of $Y=g(X)$
Discrete	$EY=E[g(X)]=\sum_x g(x)f(x)$
Continuous	$EY=E[g(X)]=\int_{-\infty}^{\infty} g(x)f(x)\,dx$

	f	F
Discrete		
Continuous		

Example: Loss to humankind of global warming

X :sustained global mean temperature increase (degrees Celsius) due to doubling of CO_2
$f(x)$:probability density, Martin Weitzman: $F(6.4) = 0.95$
$L(x)$:loss due to temperature increase due to release of CO_2

$$E[L] = \int_0^\infty L(x)\, f(x)\, dx \stackrel{?}{=} \infty$$

Simplistic model (not Weitzman's!)
$f(x) = \lambda e^{-\lambda x} \quad x > 0$
$F(x) = 1 - e^{-\lambda x} \quad x > 0; \quad \lambda = -\ln(.05)/6.4 = .47 \Leftrightarrow F(6.4) = 0.95$
$L(x) = Ae^{ax}$

$$E[L] = \int_0^\infty A\lambda e^{(a-\lambda)x} dx = \begin{matrix} \infty & a \geq \lambda \\ \frac{A\lambda}{\lambda - a} & a < \lambda \end{matrix}$$

Expectation, variance, covariance, correlation - interval scales!

$E\left(aX+b\right)=a\left(EX\right)+b$
$E\sum_{i=1}^{n}X_i=\sum_{i=1}^{n}EX_i$

$var\left(X\right)=\sigma_X^2=E\left(X-EX\right)^2=EX^2-\left(EX\right)^2$ (what is σ_X called?)
$var\left(aX+b\right)=a^2 var\left(X\right)$ (X value in NOK, $a=1/1000$ what is the unit of σ_{aX} ?)
$var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}varX_i$ when the variables are uncorrelated
$var(X+Y)=varX+2cov(X,Y)+varY$

$cov\left(X,Y\right)=E\left[\left(X-EX\right)\left(Y-EY\right)\right]$ (Y age in years X as above, unit of $cov\left(X,Y\right)$?)
$corr(X,Y)=\frac{cov(X,Y)}{\sigma_X\sigma_Y}$ (unit of $corr\left(X,Y\right)$?)

- Do SW: 2.7

Random sample of iid (independent and identically distributed) variables

$X_1, \cdots, X_n$ are iid, $EX_i = \mu$, $varX_i = \sigma^2$
$E\overline{X}_n = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$, $var\overline{X}_n = \frac{1}{n}\sigma^2$, $\sigma_{\overline{X}_n} = \frac{1}{\sqrt{n}}\sigma$

Law of large numbers:
$\overline{X}_n \xrightarrow{P} \mu$ that is $\forall c > 0$, $P\left(|\overline{X}_n - \mu| > c\right) \to 0$ as $n \to \infty$

Central limit theorem:
$\sigma < \infty \Rightarrow \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{L} N(0,1)$ that is $\forall x$,
$P\left(\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$ as $n \to \infty$

- Do SW: 2.15

Conditional distributions

Joint density: $f_{XY}(x,y) = f(x,y) \geq 0$, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$
Marginal density: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$
Conditional density: $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$
Conditional mean and variance:
$E[X|Y=y] = \mu_X(y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$,
$var[X|Y=y] = \sigma_X^2(y) = E\left[(X-\mu_X(y))^2 |Y=y\right]$

Double expectation: $EX = E[E[X|Y]] \stackrel{def}{=} E[\mu_X(Y)]$
Law of total variance:
$varX = E(var[X|Y]) + var(E[X|Y]) \stackrel{def}{=} E\left[\sigma_X^2(Y)\right] + var(\mu_X(Y))$

Example: $Y =$gender (0 for man, 1 for woman) $X =$ number of children

Example: Offsprings by gender

Brown, G.R., Laland, K.N. and Mulder, G.B. 2009. Bateman's principles and human sex roles *Trends in Ecology & Evolution* **6**:297-304

(greater mating variance in men than women; greater reproductive variance in men than women; correlation between mating and reproductive success)

Table 1. Mean and variance in reproductive success (RS) of males and females in 18 populations[a]

Country	Population or ethnic group	N_m[b]	$Mean_m$	Var_m	N_f	$Mean_f$	Var_f	V_m: V_f	I_m: I_f	Mating system[c]
Finland	1745–1900 genealogies	125	3.4	6	138	3.5	7.6	**0.79**	0.81	Monogamy
Norway	1700–1900 genealogies	955	4.7	8.5	991	4.5	8.3	**1.02**	0.98	Monogamy
Pitcairn Island	Genealogical records	145	4.6	23.6	127	4.7	23.2	**1.02**	1.04	Monogamy
Iran	Yomut Turkmen	267	5.1	8.1	216	3.9	7.1	**1.14**	0.86	Polygyny/monandry
Sweden	1825–1896 genealogies	1201	2.1	11.5	1050	2.4	9.7	**1.18**	1.65	Monogamy
Dominica	Local population	130	4.4	14.3	124	5	11.6	**1.23**	1.40	Monogamy
Tanzania	Pimbwe	138	6.0	9	154	6.1	7.3	**1.24**	1.27	Serial monogamy
USA	General social survey	1099	2.0	2.3	1344	2.0	1.8	**1.27**	1.25	Monogamy
Central African Republic	Aka	29	6.3	8.6	34	6.2	5.2	**1.66**	1.63	Polygyny/monandry
Botswana	Dobe !Kung	35	5.1	8.6	62	4.7	4.9	**1.77**	1.61	Serial monogamy
Tanzania	Hadza	54	4.3	9.8	44	3.6	5.1	**1.93**	1.63	Polygyny/serial monan
Venezuela	Yanomamo	279	3.7	10.1	380	3.4	4.4	**2.30**	2.11	Polygyny/monandry
Chad	Dazagada	44	8.6	15.0	33	6.4	6.5	**2.31**	1.72	Polygyny/monandry
Chad	Arabs	23	10.3	14.4	22	8.3	5.1	**2.82**	2.28	Polygyny/monandry
Brazil	Xavante	62	3.6	12.1	44	3.6	3.9	**3.10**	3.10	Polygyny/serial monan
Kenya	Kipsigis	82	10.9	24.4	260	6.6	5.9	**4.18**	2.52	Polygyny/monandry
Paraguay	Ache	48	6.4	15.1	25	7.8	3.6	**4.22**	5.16	Serial monogamy
Mali	Dogon	44	6.1	10.7	48	3.2	2.3	**4.75**	2.47	Polygyny/serial monan

- Do SW: 2.9