Econometria e Métodos de Modelização I

Licenciatura em Economia

REVISÃO DE ALGUNS CONCEITOS EM ESTATÍSTICA

Luís Filipe Martins

luis.martins@iscte.pt

http://home.iscte.pt/~lfsm

Departamento de Métodos Quantitativos,

ISCTE - Escola de Gestão

Lisboa, Janeiro de 2006

1 Variável Aleatória e Funções de Densidade e Probabilidade

- Uma variável aleatória X é a informação probabilistica de uma realização aleatória. O conjunto de realizações é composto pelos potenciais acontecimentos mutuamente exclusivos do processo aleatório. Uma realização é aleatória porque está associada a uma probabilidade (proporção de vezes que ocorre no longo prazo) na experiência.

Example 1 $X = \#$ *de caras quando se atira ao ar uma moeda válida por* 10 *vezes. O espaço probabilistico é* $\{0, 1, ..., 10\}$ *, uma particular realização é* $x = 6$ *e um particular evento é ”obter um número par de caras”.*

- Variável aleatória Discreta *versus* Continua.

Example 2 *Bernoulli e Poisson versus Normal and Chi-Quadrado.*

- Função densidade probabilistica discreta (pdf), $p_j = P(X = x_j) = f(x_j)$ onde a soma dos $p_j's$ é 1 e estes estão no intervalo entre 0 e 1.

- Função distribuição cumulativa (cdf), $F(x) = P(X \leq x)$. A area ou integral (caso continuo) abaixo da pdf é 1; $F(x)$ é não decrescente em relação a x and está entre 0 e 1.

Para qualquer c, $P(X \geq c) = P(X > c) = 1 - F(c)$

Para qualquer $a < b$, $P(a < X \leq b) = F(b) - F(a)$.

- Função densidade (continua) (pdf), $f(x_j) = \frac{\partial F(x)}{\partial x}|x = x_j$.
- Pdf conjunta, $f_{X,Y}(x,y) = P(X = x, Y = y)$, para X, Y v.a. discretas; e $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$.
- $\sum_{i,j} f_{X,Y}(x_j, y_i) = 1$; $\int\int f_{X,Y}(x,y)\,dxdy = 1$.
- Def: X, Y são independentes se e só se $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, para todo x, y.

Pdf Marginal , $f_X(x)$. Para v.a. discreta, $f_X(x_j) = \sum_{i=1}^{m} f_{X,Y}(x_j, y_i)$; para continua, $f_X(x) = \int f_{X,Y}(x,y)\,dy = \frac{\partial F_{X,Y}(x,\infty)}{\partial x} = \frac{\partial F_X(x)}{\partial x}$

- Pdf conditional, $f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, para todo x em que $f_X(x) > 0$.

X, Y são independentes se $f_{Y/X}(y/x) = f_Y(y)$.

- $\int f_{Y/X}(y/x)\,dy = 1$; $F_{Y/X}(y/x) = \int_{-\infty}^{y} f_{Y/X}(z/x)\,dz$.

2 Valor Esperado e Variância

- Valor esperado $E(X) \equiv \mu$ como medida de tendência central.

  $E(X) = \sum_{j=1}^{k} x_j f(x_j)$ ou $E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx$.

  Para X, Y disc., $E[g(X,Y)] = \sum_{j=1}^{k} \sum_{i=1}^{m} g(x_j, y_i) f_{X,Y}(x_j, y_i)$

- Propriedades:

  Para *c, a, b* e $\{a_1, ..., a_n\}$ qq constantes e $\{X_1, ..., X_n\}$ v.a.,

  M1: $E(c) = c$

  M2: $E(aX + b) = aE(X) + b$

  M3: $E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$

- Mediana $Med(X)$ como medida de tendência central.

- Distribuições (A)Simétricas.

- Variância $Var(X) \equiv \sigma^2$ como medida de variabilidade.

  $Var(X) = E\left[(X - \mu)^2\right] > 0$, e desvio padrão $sd(X) \equiv \sigma = +\sqrt{Var(X)}$

  Facto: $\sigma^2 = E(X^2) - \mu^2$.

- Propriedades:

  V1: $Var(X) = 0$ se e só se $P(X = c) = 1$

  V2: $Var(aX + b) = a^2 Var(X), sd(aX + b) = |a|\, sd(X)$.

- Variável aleatória Standardizada, $Z \equiv \frac{X - \mu}{\sigma}$.

3 Correlação, Independencia e Momentos Condicionados

- Covariância $Cov(X,Y) \equiv \sigma_{XY}$ e Correlação $Corr(X,Y) \equiv \rho_{XY}$ como medidas de associação (linear).

  $Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right], Corr(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

  Facto: $\sigma_{XY} = E(XY) - \mu_X \mu_Y$

- Propriedades:

  C1: Se X, Y são independentes, então $Cov(X,Y) = 0$.

  C2: $Cov\left(a_1 X + b_1, a_2 Y + b_2\right) = a_1 a_2 Cov(X,Y)$

  C3: $|\sigma_{XY}| \leq \sigma_X \sigma_Y$ (desigualdade de Cauchy-Schwartz)

  C4: $-1 \leq Corr(X,Y) \leq 1$

  C5: $Corr\left(a_1 X + b_1, a_2 Y + b_2\right) = Corr\left(X,Y\right), a_1 a_2 > 0;$

  $Corr\left(a_1 X + b_1, a_2 Y + b_2\right) = -Corr\left(X,Y\right), a_1 a_2 < 0.$

  C6: $Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X,Y)$

  C7: Se $\{X_1, ..., X_n\}$ são v.a. não autocorrelacionadas mutuamente, então $Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var\left(X_i\right).$ Caso contrário, $Var\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 Var\left(X_i\right) + 2\sum_{i>j} a_i a_j Cov\left(X_i, X_j\right)$

- Valor esperado condicional de Y dado X, $E\left(Y/x\right) = \sum_{j=1}^m y_j f_{Y/X}\left(y_j/x\right)$ ou $E\left(Y/x\right) = \int_{-\infty}^{+\infty} y f_{Y/X}\left(y/x\right) dy$, é uma função de x.

- Propriedades:

  Considere $g(X)$ e $h(X)$ qualquer duas funções.

CE1: $E\left[g\left(X\right)/X\right] = g\left(X\right)$

CE2: $E\left[g\left(X\right)Y + h(X)/X\right] = g\left(X\right)E\left(Y/X\right) + h(X)$

CE3: Se Y, X são independentes, então $E\left(Y/X\right) = E\left(Y\right)$. Se $E\left(Y/X\right) = E\left(Y\right)$, então $Cov(g\left(X\right), Y) = 0.$

CE4: $E_X\left[E\left(Y/X\right)\right] = E\left(Y\right)$ (lei das expectativas iteradas)

- Variância condicional de Y dado X, $Var\left(Y/x\right)$

$$
\begin{aligned}
Var\left(Y/X=x\right) &= E\left(\left[Y - E\left(Y/X\right)\right]^2/x\right) \\
&= \sum_{j=1}^{m}\left[y_j - E\left(Y/x\right)\right]^2 f_{Y/X}\left(y_j/x\right) \\
Var\left(Y/X=x\right) &= E(Y^2/x) - \left[E\left(Y/X\right)\right]^2
\end{aligned}
$$

Se Y, X são independentes, então $Var\left(Y/X\right) = Var\left(Y\right).$

4 Principais Distribuições

- Distribuição Normal, $X \sim N\left(\mu, \sigma^2\right)$ é simétrica em torno de μ e $0.95 = P\left(\mu - 1.96\sigma < X < \mu - 1.96\sigma\right)$

  Tabelas: Normal Standard, $Z = \frac{X-\mu}{\sigma}$; $P\left(Z \leq z\right) = \Phi\left(z\right)$

  $P\left(Z > z\right) = 1 - \Phi\left(z\right)$ e $P\left(a \leq Z \leq b\right) = \Phi\left(b\right) - \Phi\left(a\right)$

  Qualquer combinação linear de $i.i.d.N.$ tem uma distribuição $N.$

Example 3 *Para* $Y_i \sim N\left(\mu, \sigma^2\right), i = 1, ..., n,$ *então* $\frac{1}{n}\sum_{i=1}^{n} Y_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$

- Normal Bivariada, $N_2\left(\boldsymbol{\mu}, \Omega\right)$

$$\begin{aligned} & f_{X,Y}\left(x, y\right) \\ = & \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \exp\left[-\frac{1}{2}\left(\frac{Z_X^2 - 2\rho_{XY}Z_XZ_Y + Z_Y^2}{1-\rho_{XY}^2}\right)\right], \end{aligned}$$

$$Z_X = \frac{x-\mu_X}{\sigma_X}; Z_Y = \frac{y-\mu_Y}{\sigma_Y}.$$

  Se X, Y são $N_2\left(\boldsymbol{\mu}, \Omega\right)$, então X, Y são independentes se e só se $Cov(X, Y) = 0$.

- Distribuição do Chi-Quadrado com n graus de liberdade, $X \sim \chi^2_{(n)}$ é assimétrica e X é não negativa c.p.1

  $E(X) = n, Var(X) = 2n$

Se $Z_i \sim i.i.d.N\left(0,1\right), i = 1, ..., n$, então $X = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$

- Distribuição do t-Student com n graus de liberdade, $X \sim t_{(n)}$ é simétrica em torno de zero.

$E(X) = 0, n > 1, Var(X) = \frac{n}{n-2}, n > 2$ (para que exista - finito).

$t_{(n)}$ tem mais massa nas abas do que a $N\left(0,1\right)$ e $t_{(\infty)}$ é igual a $N\left(0,1\right)$ c.p.1

Se $Z \sim N\left(0,1\right), X \sim \chi^2_{(n)}, Z, X$ independentes, então $T = \frac{Z}{\sqrt{\frac{X}{n}}} \sim t_{(n)}$.

- Distribuição do F-Snedcor com (k_1, k_2) graus de liberdade, $X \sim F_{(k_1,k_2)}$ é assimétrica e X é não negativa c.p.1

Se $X_1 \sim \chi^2_{(k_1)}, X_2 \sim \chi^2_{(k_2)}, X_1, X_2$ independentes, então $F = \frac{X_1/k_1}{X_2/k_2} \sim F_{(k_1,k_2)}$.

5 Estimação

- Processo de ”aprendizagem” sobre a população (desconhecido) com o recurso à informação contida numa amostra disponível que é extraída da população. Estimação Pontual; Estimação de Intervalos e Ensaios de Hipóteses. Exemplo de um objecto de interesse: O valor esperado da população.
- Suponha que a pdf dos $Y's$ é dada por $f(y;\theta)$. Então, $\{Y_1,...,Y_n\}$ é uma amostra aleatória de $f(y;\theta)$ se $Y_i \sim i.i.d. f(y;\theta), i = 1,...,n$. Os dados $\{y_1,...,y_n\}$ são habitualmente diferentes para cada amostra distinta.
- Estimador Pontual para θ, $W = h(Y_1,...,Y_n)$ e estimativa $w = h(y_1,...,y_n)$
- Propriedade de Amostra Finita da distribuição da v.a. W:

W é centrado se $E(W) = \theta$. $Bias(W) = E(W) - \theta$.

$Var(W)$ (eficiência)

W_1 é eficiente relativamente a W_2 quando $Var(W_1) < Var(W_2)$, com W_1, W_2 centrados.

Mean squared error (erro quadrático médio) $MSE(W) = Var(W) + [Bias(W)]^2$.

Example 4 *A média amostral $\overline{Y}$ é um estimador para o valor esperado (média populacional) μ e a variância amostral $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2$ para a variância (populacional) σ^2.*

6 Intervalo de Confiança

- Estimação de Intervalos e Intervalo de Confiança para o exemplo $Y \sim N(\mu, 1)$. Neste caso, $\overline{Y} \sim N\left(\mu, \frac{1}{n}\right)$.

  (IC) $P\left(\overline{Y} - \frac{1.96}{\sqrt{n}} < \mu < \overline{Y} + \frac{1.96}{\sqrt{n}}\right) = 0.95$

  (IE) $\overline{Y} - \frac{1.96}{\sqrt{n}} < \mu < \overline{Y} + \frac{1.96}{\sqrt{n}}$ é um intervalo aleatório

  $P\left(\overline{y} - \frac{1.96}{\sqrt{n}} < \mu < \overline{y} + \frac{1.96}{\sqrt{n}}\right) = 0$ ou 1

- Outros resultados importantes para o parâmetro de interesse μ:

  (σ conhecido) $\left[\overline{y} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{y} + 1.96\frac{\sigma}{\sqrt{n}}\right]$ porque $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

  (σ desconhecido, $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2}$). Porque $\frac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$ então, o IC a $100(1-\alpha)\%$ é dado por $\left[\overline{y} - c_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}}, \overline{y} + c_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}}\right]$. Lembrar que para $\alpha = 0.05$, $c_{(n-1,\alpha/2)}$ tende para 1.96 (aprox. 2) quando n tende para infinito.

  Quando, Y NÃO é normalmente distribuido, desde que n é suficientemente elevado, pelo TLC (CLT), pode-se usar a APROXIMAÇÃO a 95% IC $\left[\overline{y} - 1.96\frac{s}{\sqrt{n}}, \overline{y} + 1.96\frac{s}{\sqrt{n}}\right]$.

- Em relação ao parâmetro de interesse σ^2, usar o resultado $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ onde $Y \sim N\left(\mu, \sigma^2\right)$.

7 Inferência

- Passos num Ensaio de Hipóteses:

(1) Definir a hipótese nula H_0 e hipótese alternativa H_1 em relação ao parâmetro (por exemplo) de interesse, que é desconhecido e do qual se procura inferir (ex: $H_0 : \mu = 1$ *vs* $H_1 : \mu \neq 1$)

(2) Selecionar a estatistica de teste apropriada T (ex: $T = \overline{Y}$)

(3) Determinar a distribuição da estatistica de teste T (ex: $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$)

(4) Escolher o nível de significância do teste $\alpha = P\left(\operatorname{Re} j H_0 / H_0\right)$ (erro do tipo I). Default: $\alpha = 0.05$. Com base em α, e sob H_0, obter o(s) valor(es) critico(s) vc :

  (a) H_1 Unilateral: $P_{H_0}(T > vc) = \alpha$, para $H_1 : \theta > \theta_0$, OU $P_{H_0}(T < vc) = \alpha$, para $H_1 : \theta < \theta_0$.

  (b) $H_1 : \theta \neq \theta_0$ Bilateral (normalmente): $P_{H_0}(|T| > vc) = \alpha$, se simétrica OU $P_{H_0}(T > vc_1) = \frac{\alpha}{2}, P_{H_0}(T < vc_2) = \frac{\alpha}{2}$ se não é simétrica.

(5) Definir a regra de decisão (intervalos para T_{obs} delimitados com base no(s) vc) e as correspondentes regiões crítica e de aceitação. Decidir em relação ao teste vendo em que região ”cai” T_{obs}. Este procedimento é equivalente à análise do IC a $100\,(1-\alpha)\,\%$ da distribuição em 3., sob H_0.

- Potência do teste $\pi(\theta) = 1 - P_{H_1}\left(AcH_0\right)$
- $p - value$ é o maior valor para α para o qual ainda não se rejeita H_0 :

(a) H_1 Unilateral: $p = P_{H_0}(T > T_{obs})$, para $H_1 : \theta > \theta_0$.

(b) $H_1 : \theta \neq \theta_0$ Bilateral: $p = P_{H_0}(|T| > |T_{obs}|)$ se simétrica.