Capítulo 1

Introdução

1.1 Revisões

Convenções e notação: Seja X uma v.a. (discreta ou contínua), com função de distribuição $F_X(x|\theta)$, onde θ designa o parâmetro (uni ou multivariado)
Funções de v.a.: Seja X uma v.a. contínua com f.d.p. $f_X(x|\theta)$, onde θ é o parâmetro associado, que toma valores em Ω_X, e seja $Y = g(X)$, onde $g(.) : \Omega_X \to \Omega_Y$ é uma função que admite derivada contínua em todo o domínio, anulando-se quando muito num número finito de pontos. Então para cada $y \in \Omega_Y$ existe um número positivo n e números reais $v_1(y), v_2(y), \ldots, v_n(y)$ tais que

$$g(v_i(y)) = y, \quad \frac{dg(v_i(y))}{dy} \neq 0, \quad j = 1, 2, \ldots, n(y).$$

Então

$$f_Y(y|\theta) = \sum_{j=1}^{n} f_X(v_j(y)|\theta) \left| \frac{dv_j(y)}{dy} \right|, \quad n > 0, \quad f_Y(y|\theta) = 0, \quad n = 0.$$

Um caso particular ocorre quando $g(.)$ é uma função monótona (crescente ou decrescente). Nesse caso

$$f_Y(y|\theta) = f_X(g^{-1}(y)|\theta) \left| \frac{d}{dy} g^{-1}(y) \right| I_{\Omega_Y}(y).$$

Uma das transformações mais importantes é a *transformação do integral de probabilidade.*

Teorema 1 *Seja X uma v.a. contínua com f.d. $F_X(.|\theta)$, e defina-se $Y = F_X(X|\theta)$. Então Y é uniformemente distribuiída em $(0,1)$, i.e.,*

$$P(Y \leq y) = y I_{(0,1)}(y).$$

O teorema anterior é particularmente útil na geração de amostras aleatórias. Por exemplo, se for necessário gerar uma observação x de uma população com f.d. $F_X(.)$, então gera-se um número aleatório y, entre 0 e 1, e resolve-se a equação $F_X(x) = y$ (o que por si só pode ser uma tarefa hercúlea!).

Função geradora de momentos: Seja X uma v.a. (contínua ou discreta) com f.d.p. $f_X(x|\theta)$. Para $|s| \leq 1$

$$M_X(s) = E[e^{sX}].$$

Quando existe este valor esperado numa vizinhança de $s = 0$, $M_X(s)$ representa a função geradora de momentos da v.a. X, uma vez que

$$M_X^{(r)}(s) = E[X^r], \quad r \in \mathbb{N}.$$

Propriedades da f.g.m.:

(a) $M_{X-\lambda}(s) = e^{-\lambda s} M_X(s)$

(b) Se X e Y são v.a. independentes, com f.g.m. $M_X(.)$ e $M_Y(.)$ respectivamente, então existe f.g.m. de $X+Y$ e é dada pela expressão:

$$M_{X+Y}(s) = M_X(s) M_Y(s)$$

(c) A f.g.m., quando existe, determina univocamente a função de distribuição. Reciprocamente, a f.g.m., se existir, é única.

(d) Se a f.g.m. do par aleatório (X, Y) existir, o par (X, Y) possui momentos de todas as ordens, os quais podem obter-se da relação

$$\left. \frac{\delta^{r+k} M_{(X,Y)}(s_1, s_2)}{\delta s_1^r \delta s_2^k} \right|_{s_1 = s_2 = 0} = E[X^r Y^k]$$

(e) As v.a. X e Y são independentes se e só se $M_{X+Y}(s_1, s_2) = M_{(X,Y)}(s_1, 0) M_{(0,Y)}(0, s_2)$.

1.2 Variáveis aleatórias

1.2.1 Discretas

a) **Uniforme discreta**: diz-se que X tem distribuição uniforme discreta em $\Xi = \{x_1, x_2, \ldots, x_n\}$ se

$$f_X(x|n) = \frac{1}{n} I_\Xi(x).$$

Caso particular: $\Xi = \{0, 1, \ldots, n\}$. Nesse caso

$$E[X] = \frac{n+1}{2}, \quad Var[X] = \frac{n^2 - 1}{12}, \quad M_X(s) = e^s \frac{1 - e^{sn}}{1 - e^s}.$$

b) **Binomial**: $X \sim Bin(n, p)$, descreve os processos em que se realizam repetidas provas independentes de Bernoulli, sendo p a probabilidade (constante) de sucesso. A f.d.p. é dada por

$$f_X(x|n, p) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{1,2,\ldots,n\}}(x)$$

sendo que

$$E[X] = np, \quad Var[X] = np(1-p), \quad M_X(s) = e^{-nsp}(1 - p + pe^s)^n.$$

Propriedade relevante:

- Se as v.a. $X_i, i = 1, \ldots, k$ são independentes, então

$$X_i \sim Bin(n_i, p) \Rightarrow \sum_{i=1}^{k} X_i \sim Bin(\sum_{i=1}^{k} n_i, p).$$

d) **Binomial negativa**: $X \sim BN(r, p)$, onde X designa o número de provas de Bernoulli independentes que têm de realizar-se até se obterem r sucessos, sendo p a probabilidade de sucesso (constante ao longo das repetições). A f.d.p. é dada por

$$f_X(x|r,p) = \binom{x-1}{x-r} p^r (1-p)^{x-r} I_{\{r,r+1,\ldots\}}(x)$$

com

$$E[X] = \frac{r}{p}, \quad Var[X] = r\frac{1-p}{p^2}, \quad M_X(s) = (pe^s)^r (1-(1-p)e^s)^{-r}.$$

Se em vez de X se modelar $Y = X - R$ (nº de insucessos observados até ocorrer o r-ésimo sucesso), então a f.d.p. é dada por

$$f_Y(y|r,p) = \binom{r+y-1}{y} p^r (1-p)^y I_{\{0,1,\ldots\}}(y)$$

com

$$E[Y] = r\frac{1-p}{p}, \quad Var[Y] = r\frac{1-p}{p^2}, \quad M_X(s) = p^r (1-(1-p)e^s)^{-r}.$$

Propriedades relevantes:

- Se X_i são v.a. independentes (para $i = 1, 2, \ldots, k$) então

$$X_i \sim BN(r_i, p) \Longrightarrow \sum_{i=1}^{n} X_i \sim BN(\sum_{i=1}^{n} r_i, p).$$

- Se $X \sim BN(1, p)$ então diz-se que X tem distribuição geométrica de parâmetro p (simbolicamente: $X \sim Geo(p)$)
- Se $X \sim Geo(p)$ então para $m, n \in I\!N : P(X > m + n | X > m) = P(X \geq n)$ (propriedade da amnésia da geométrica)
- Se X assume valores da sucessão $0, 1, \ldots$, e $\forall m, n \in I\!N : P(X > m + n | X > m) = P(X \geq n)$, então X tem distribuição Geométrica.

(e) **Poisson**: $X \sim Poisson(\lambda)$ com f.d.p. dada por

$$f_X(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} I_{I\!N_0}(x)$$

com

$$E[X] = Var[X] = \lambda, \quad M_X(s) = e^{\lambda(e^s - 1)}.$$

Propriedades mais relevantes:

- Se as v.a. $X_i, i = 1, \ldots, k$ são independentes, então

$$X_i \sim Poisson(\lambda_i) \Leftrightarrow \sum_{i=1}^{k} X_i \sim Poisson(\sum_{i=1}^{k} \lambda_i)$$

- Se $X \sim Poisson(\lambda)$ e $Y|X = x \sim Bin(x, p)$ então $Y \sim Poisson(\lambda, p)$.
- A distribuição binomial de parâmetros (n, p) converge para a de Poisson quando $n \to \infty$ e $p \to 0$, mantendo-se constante o produto np.
- A família de distribuições Binomial Negativa inclui a distribuição de Poisson como caso limite, fazendo $r \to \infty$ tal que $r(1-p) \to \lambda$, com $\lambda \in (0, \infty)$.

(f) **Hipergeométrica**: $X \sim HG(N, n, M)$, onde X designa o número de sucessos em n provas de Bernoulli extraídas sem reposição de uma população de dimensão N onde há M sucessos.

$$f_X(x|N, n, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} I_{[\max(0, n-M), \min(n, M)]}(x)$$

com

$$E[X] = n\frac{M}{N}, \quad Var[X] = n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}$$

Propriedades mais relevantes:

- Com n e M fixos:

$$\lim_{N\to\infty}\left[\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}\right] = \binom{n}{x} p^x (1-p)^{n-x}$$

  com $p = \frac{M}{N}$. Assim sendo a f.d.p. de uma hipergeométrica tende para a f.d.p. de uma binomial quando a dimensão da população aumenta, mantendo-se os outros parâmetros constantes.
- A variância associada à distribuição hipergeométrica é inferior à da binomial correspondente. Esta propriedade pode ser, nalguns casos, uma vantagem da distribuição Hipergeométrica em relação à Binomial. No entanto, e de acordo com a propriedade anterior, à medida qua a dimensão da população aumenta, esbate-se a diferença entre extracções com reposição e sem reposição e, consequentemente, a diferença entre as duas distribuições.

(g) **Multinomial**: $X \sim MN(n, \mathbf{p})$, com $\mathbf{p} = (p_0, p_1, p_2, \ldots, p_k)$. A Multinomial generaliza a binomial: trata-se de uma sucessão de provas independentes, em que em cada prova realiza-se um e um só dos $k+1$ acontecimentos de uma partição do espaço amostral $\{A_0, A_1, \ldots, A_k)$. Seja $p_i = P(A_i)$, com $p_0 + p_1 + \ldots + p_k = 1$. À semelhança da binomial, interessa estudar a probabilidade de em n provas observar x_0 vezes o acontecimento A_0, x_1 vezes o acontecimento A_1, etc.

$$f_X(\mathbf{x}|n, \mathbf{p}) = \frac{n!}{x_0!x_1!\ldots, x_k!}\prod_{i=0}^{k} p_i^{x_i}, \quad \mathbf{x} = (x_0, x_1, \ldots, x_k)$$

$$E[X_i] = np_i, \quad Var[X_i] = np_i(1-p_i), \quad Cov(X_i, X_j) = -np_ip_j$$

e, finalmente,

$$M_X(\mathbf{s}) = \left(\sum_{i=0}^{k} p_i e^{s_i}\right)^n$$

1.2.2 Contínuas

(a) Uniforme contínua $X \sim Unif(a,b)$, com $a < b$ e

$$f_X(x|a,b) = \frac{1}{b-a} I_{[a,b]}(x)$$

com

$$E[X] = \frac{a+b}{2}, \quad Var[X] = \frac{(b-a)^2}{12}$$

(b) Gamma $X \sim Gama(\alpha, \beta)$, para $\alpha, \beta \in I\!R^+$, onde α se designa por *parâmetro de forma* e β é o *parâmetro de escala.*

$$f_X(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{I\!R^+}(x)$$

com $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$.

$$E[X] = \frac{\alpha}{\beta}, \quad Var[X] = \frac{\alpha}{\beta^2}, \quad M_X(s) = \left(1 - \frac{s}{\beta}\right)^{-\alpha} I_{(0,\beta)}(s).$$

Propriedades mais relevantes:

- Se $X \sim Gama(\alpha, \beta)$ e se $\alpha \in I\!N$, então para qualquer x:

  $$P(X \leq x) = P(Y \geq \alpha)$$

  com $Y \sim Poisson(\beta x)$. Nesse caso X diz-se ter distribuição *Erlang*.
- Se $\alpha = \frac{p}{2}$, sendo p um inteiro, e $\beta = \frac{1}{2}$, então diz-se que $X \sim \chi^2_{(p)}$ (qui-quadrado com p graus de liberdade).
- Se $\alpha = 1$ então $X \sim Exp(\beta)$ (exponencial com parâmetro β).
- A distribuição exponencial tem a propriedade da amnésia, i.e.

  $$P(X > s + t | X > t) = P(X > s), \quad \forall s, t \in I\!R^+$$

- Se $X \sim Exp(\beta)$ então $Y = X^{1/\gamma} \sim Weibull(\gamma, \beta)$ (Weibull de parâmetros γ e β), a que corresponde a f.d.p.

  $$f_Y(y|\gamma,\beta) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta}, \quad y, \gamma, \beta \in I\!R^+.$$

(c) Normal $X \sim N(\mu, \sigma)$, $\mu \in I\!R, \sigma \in I\!R^+$, com

$$f_X(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in I\!R$$

com

$$E[X] = \mu, \quad Var[X] = \sigma^2, \quad M_X(s) = e^{\mu + \frac{\sigma^2 s^2}{2}}$$

Propriedades mais relevantes

- Se X_i forem v.a. independentes então

  $$X_i \sim N(\mu_i, \sigma_i) \Rightarrow \sum_{i=1}^k \alpha_i X_i \sim N(\sum_{i=1}^k \alpha_i \mu_i, \sqrt{\sum_{i=1}^k \alpha_i^2 \sigma_i^2})$$

- Se X e Y forem v.a. independentes e se $X+Y$ tem distribuição normal, então também X e Y possuem distribuição normal.

(d) Beta $X \sim Beta(\alpha, \beta)$, com

$$f_X(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1} I_{(0,1)}(x)$$

com $\alpha, \beta \in I\!R^+$, onde

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

a chamada *função beta*, a qual se relaciona com a função gama através da seguinte relação:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Tem-se

$$E[X] = \frac{\alpha}{\alpha+\beta}, \quad Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

e

$$M_X(s) = \sum_{k=0}^{\infty} \frac{\Gamma(\alpha+k)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+k)\Gamma(\alpha)} \frac{s^k}{\Gamma(k+1)}$$

a qual se revela pouco importante nas aplicações.

Propriedades mais importantes

- Se X e Y são v.a. independentes, com $X \sim Gama(m, \beta)$ e $Y \sim Gama(n, \beta)$ então $\frac{X}{X+Y} \sim Beta(m, n)$.
- A forma da f.d.p. de uma beta varia muito, consoante os parâmetros. Por exemplo, uma $Beta(\alpha, \beta)$ com $\alpha > 1$ e $\beta = 1$ é estritamente crescente, se $\alpha = 1$ e $\beta > 1$ é estritamente decrescente, em forma de U se $\alpha < 1$ e $\beta < 1$, etc.

(e) Cauchy $X \sim Cauchy(\lambda, \theta)$, com $\theta \in I\!R$ e $\lambda > 0$, com

$$f_X(x|\lambda, \theta) = \frac{\lambda}{\pi[\lambda^2 + (x-\theta)^2]}, \quad x \in I\!R$$

Não tem momentos de ordem alguma, nem f.g.m. O parâmetro θ descreve o ponto de simetria da f.d.p. (é a mediana).

Propriedades mais relevantes:

- Se as v.a. X_i $(i = 1, 2, \ldots)$ são independentes, então

$$X_i \sim Cauchy(\lambda_i, \theta_i) \Rightarrow \sum_{i=1}^{n} X_i \sim Cauchy(\sum_{i=1}^{n} \lambda_i, \sum_{i=1}^{n} \theta_i)$$

- Se X e Y são v.a. independentes, ambas com distribuição $N(0, 1)$ então

$$\frac{X}{Y} \sim Cauchy(0, 1)$$

(f) Outras distribuições menos comuns:

- Lognormal: se X é uma v.a. positiva e se $lnX \sim N(0,1)$, então diz-se que X tem distribuição lognormal, com

$$f_X(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\{-\frac{(lnx-\mu)^2}{2\sigma^2}\}$$

com $E[X] = e^{\mu+\sigma^2/2}$ e $Var[X] = e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$.

- Pareto: corresponde à f.d.p.

$$f_X(x|\theta,x_0) = \frac{\theta}{x_0}\left(\frac{x_0}{x}\right)^{\theta+1}, \quad x > x_0 > 0, \theta > 0$$

com $E[X] = \frac{\theta x_0}{\theta-1}$, se $\theta > 1$, e $Var[X] = \frac{\theta x_0^2}{\theta-2} - \left(\frac{\theta x_0}{\theta-1}\right)^2$, se $\theta > 2$.

- Laplace: corresponde à f.d.p.

$$f_X(x|\lambda,\theta) = \frac{\lambda}{2}\exp\{-\lambda|x-\theta|\}, \quad x,\theta \in I\!R, \lambda > 0$$

com $E[X] = \theta$ e $Var[X] = \frac{2}{\lambda^2}$.

- Logística: tem f.d.p.

$$f_X(x|\alpha,\beta) = \frac{e^{-(x-\alpha)/\beta}}{\beta[1+e^{-(x-\alpha)/\beta}]^2}, \quad x \in I\!R, \alpha \in I\!R, \beta > 0$$

com $E[X] = \alpha, Var[X] = \pi^2\beta^2/3$.

1.3 Família Exponencial

Muitas das distribuições dadas até aqui enquadram-se num conjunto de famílias de distribuições designado por **família exponencial**.

Uma família de f.d.p. $\{f(.|\theta), \theta \in \Theta\}$ pertence à família exponencial se a sua f.d.p. poder ser escrita da seguinte forma:

$$f(x|\theta) = h(x)c(\theta)e^{\sum_{i=1}^k w_i(\theta)t_i(x)}$$

com $h(x) \geq 0, \forall x$, $t_1, \ldots, t_k$ são funções reais da observação x (e que, em particular, não dependem de θ), $c(\theta) \geq 0$ e $w_1(\theta), \ldots, w_k(\theta)$ são funções reais de θ (e que, em particular, não dependem de x).

Por abuso de linguagem, sempre que dissermos *a v.a. X pertence à família exponencial* significa que *a f.d.p. da v.a. X pertence à família exponencial*.

Mais tarde veremos a importância da família exponencial. Para além das suas propriedades matemáticas, a sua forma implica propriedades estatísticas que tornam o seu tratamento especialmente simples.

Para já note-se que se tivermos uma série longa de observações de uma v.a. X cuja f.d.p. pertence à família exponencial, então só necessitamos de calcular k parcelas, funções de todos os dados da amostra. Isto significa em particular que se tivermos duas amostras $\underline{x}$ e $\underline{y}$, com

$$\underline{x} = (x_1, \ldots, x_{n_1}), \quad \underline{y} = (y_1, \ldots, y_{n_2})$$

tais que

$$\prod_{i=1}^{n_1} h(x_i) = \prod_{j=1}^{n_2} h(y_j$$

e

$$\forall r = 1, 2, \ldots, k : \sum_{i=1}^{n_1} t_r(x_i) = \sum_{j=1}^{n_2} t_w(y_j), \quad \text{para algum } w = 1, 2, \ldots, k$$

então $f(\underline{x}|\theta) = f(\underline{y}|\theta)$.

Como veremos mais tarde esta propriedade permitirá a redução dos dados amostrais, levando à consideração de estatísticas suficientes (as quais serão tema do próximo capítulo). Note-se ainda que se o conjunto de valores de x para os quais $f_X(x|\theta) > 0$ depender de θ, então $f_X(.|\theta)$ não pertence à família exponencial. Por exemplo, se $X \sim Unif(a,b)$, então X não pertence à família exponencial.

Por vezes a família exponencial é reparametrizada da seguinte forma:

$$f(x|\eta) = h(x)c^\star(\eta)e^{\sum_{i=1}^k \eta_i t_i(x)}$$

, tal que $\eta_i = w_i(\theta)$. O conjunto $H = \{\eta = (\eta_1, \ldots, \eta_k) : \int_{-\infty}^{\infty} h(x)e^{\sum_{i=1}^k \eta_i t_i(x)}dx < \infty\}$ é designado por *espaço natural paramétrico*. Em particular $\eta = (w_1(\theta), \ldots, w_k(\theta))$ pertence ao espaço paramétrico natural, que pode conter mais elementos para além deste. A vantagem de considerar o espaço paramétrico natural é que se trata de um conjunto convexo; a desvantagem é que a sua interpretação é geralmente mais difícil que a do espaço original.

1.4 Famílias de localização e escala

Quando normalizamos a $N(\mu, \sigma^2)$, passando-a a $N(0,1)$, dizemos que estamos a fazer uma transformação de localização (ao subtrair μ, o valor esperado), seguida de uma transformação de escala (ao dividir por σ). Este procedimento pode ser alargado a mais f.d.p. A f.d.p. que resulta de uma transformação de localização e escala chama-se, à imagem do caso normal, uma f.d.p. *padrão* ou em standard.

Teorema 2 *Seja $f(.)$ uma f.d.p. arbitrária, e $\mu, \sigma > 0$ duas constantes quaisquer. Então a função*

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$$

é uma f.d.p.

Este teorema permite-nos então definir uma família de localização e escala.

DEfinição 3 *Seja $f(.)$ uma f.d.p. arbitrária. Então $\forall \mu \in I\!R, \sigma \in I\!R^+$, a família de f.d.p. $\{\frac{1}{\sigma} f(\frac{x-\mu}{\sigma}, (\mu, \sigma) \in (I\!R, I\!R^+)\}$ é uma família de localização e escala com f.d.p. padrão $f(.)$. Neste caso μ diz-se ser o parâmtero de localização e σ o parâmetro de escala.*

O efeito da introdução do parâmetro de licalização e de escala serve para esticar (se $\sigma > 1$ ou contrair (se $\sigma < 1$) o gráfico da f.d.p., sendo que a introdução do parâmetro de localização serve para transladar o gráfico.

Os teoremas que se seguem mostram a importância das famílias de localização e escala para determinadas transformações.

Teorema 4 *Seja* $f(.)$ *uma função arbitrária,* $\mu \in I\!R, \sigma \in I\!R^+$ *também arbitrários. Então existe uma v.a.* X *com f.d.p.* $\frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ *sse existe uma v.a.* Z *com f.d.p.* $f(.)$ *tal que*

$$X = \sigma Z + \mu.$$

Teorema 5 *Seja* Z *uma v.a. com f.d.p.* $f(.)$ *tal que* $E[Z]$ *e* $Var[Z]$ *existem. Se* X *for uma v.a. com f.d.p.* $\frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ *então*

$$E[X] = \sigma E[Z] + \mu, \quad Var[Z] = \sigma^2 Var[X].$$

Note-se que se $\sigma = 1$ então teremos simplesmente uma família de localização, enqaunto que se $\mu = 0$ teremos uma família de escala.

Famílias de localização e escala podem ter aplicações muito concretas. O exemplo que se segue é construído para uma família de escala.

Exemplo 6 *Seja* X *o peso se um peito de frango comercializado por uma empresa Holandesa de produção de aves.*

Informações da organização europeia de comércio alimentar indicam que uma percentagem desse peso é conseguida pela injecção de água, sendo um valor constante e igual a 35%. Assim o verdadeiro peso é $0.65X$ *(e não* X*), pelo que se* $f_X(x)$ *for a f.d.p. de* X *no ponto* x*, a f.d.p. do verdadeiro peso é* $f_X(0.65x)$*.*