Famílias de Distribuições

Introdução

(Casella e Berger pág 77)

⇒ Distribuições estatísticas são utilizadas para modelar populações.

⇒ Em geral, lidamos com uma família de distribuições em vez de uma única.

⇒ Esta família é indexada por um ou mais parâmetros, o que nos permite variar certas características da distribuição, ao mesmo tempo em que permanece com uma forma funcional.

⇒ Por exemplo, podemos especificar que a distribuição normal é uma opção razoável para modelar uma população em particular, mas não podemos especificar precisamente a média. Então, lidamos com uma família paramétrica de distribuições normais com média μ, onde μ é um parâmetro desconhecido, $-\infty < \mu < \infty$.

⇒ A: Família locação-escala (distribuições contínuas).

B: Família exponencial (distribuições discretas e contínuas).

Notação:

Distribuições discretas
Distribuições contínuas ⇒ Θ: espaço paramétrico

$\mathcal{X}$: espaço amostral ou suporte da distribuição

O conjunto Θ em que $\underset{\sim}{\theta}$ toma valores é denominado espaço paramétrico.

Modelo Normal

$$\Theta = \{ \underset{\sim}{\theta} = (\mu, \sigma^2), \ \mu \in \mathbb{R} \text{ e } \sigma^2 > 0 \}$$

$$\mathcal{X} = \{ x, \ x \in \mathbb{R} \}$$

Famílias de Locação e de Escala

Casella e Berger pág 104

→ Famílias comuns de distribuições contínuas.

→ Três técnicas para a construção de famílias de distribuições.

→ Famílias resultantes úteis na modelagem.

Os três tipos de famílias são chamadas de locação, de escala e de locação-escala.

Cada uma das famílias é construída pela especificação de uma única fdp, seja $f(x)$, chamada de fdp padrão para a família. Então, todas as outras fdps na família são geradas pela transformação da fdp padrão de um modo prescrito.

Começamos com um teorema simples sobre fdps.

Teorema Seja $f(x)$ qualquer fdp e que μ e $\sigma > 0$ sejam constantes dadas. Então, a função

$$g(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \text{ é uma fdp.}$$

Prova: Devemos verificar se $g(x|\mu,\sigma)$ é não negativa e se a integral é igual a 1.

Note que $g(x \mid \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \geq 0$, pois

$\sigma > 0$ e $f(x)$ é uma fdp, então $f(x) \geq 0$

para todos os valores de x. Assim,

$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \geq 0$ para todos os valores de

x, μ e $\sigma > 0$.

Em seguida, note que

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx = \int_{-\infty}^{\infty} f(y)\, dy = 1,$$

pois $f(y)$ é uma fdp, onde $y = \frac{x-\mu}{\sigma} \Rightarrow$

$dy = \frac{1}{\sigma} dx$.

Agora, recorremos à primeira de nossas construções, a das famílias de locação.

Definição: Seja $f(x)$ qualquer fdp. Então, a família de fdps $f(x-\mu)$, indexada pelo parâmetro μ, $-\infty < \mu < \infty$, é chamada de família locação com fdp padrão $f(x)$ e μ é chamado de parâmetro de locação para a família

<u>Comentário 01</u>: * O apontado pela definição acima é que podemos começar com qualquer fdp $f(x)$ e gerar uma família de fdp's introduzindo um parâmetro de locação.

* O efeito de introduzir o parâmetro de locação μ é que, por exemplo, o ponto que estava acima de 0 fica agora acima de μ, de modo que o formato do gráfico permanece inalterado.

<u>EXEMPLO 1</u>: <u>Família normal</u>

Por exemplo, se $\sigma>0$ é um número específico conhecido e definimos

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}x^2\right\}, \quad -\infty<x<\infty,$$

então, a família de locação com fdp padrão $f(x)$ é o conjunto de distribuições normais com média μ desconhecida e variância σ^2 conhecida. Basta substituir x por $x-\mu$ na fórmula acima e teremos

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad -\infty<x<\infty$$

e

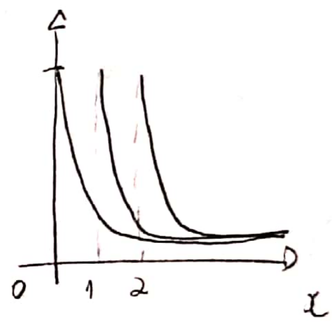
$$-\infty<\mu<\infty.$$

$\Rightarrow$ GRÁFICO no $\mathbb{R}$.

EXEMPLO 2: Família exponencial
Seja $f(x) = e^{-x}$, $x \geq 0$ e $f(x) = 0$, $x < 0$. Para formar uma família de locação, substituímos x por $x - r$ para obter

$$f(x|r) = \begin{cases} e^{-(x-r)} & x - r \geq 0 \\ 0 & x - r < 0 \end{cases}$$

$$= \begin{cases} e^{-(x-r)} & x \geq r \\ 0 & x < r \end{cases}$$



$\Rightarrow$ GRÁFICO no $\mathbb{R}$ (Gráficos de $f(x|r)$ para diversos valores de r).

Note que o formato do gráfico não se altera, mas a parte positiva do gráfico começa em r em vez de em 0.

Além disso, note que se o conjunto de x para o qual $f(x) > 0$ não é a reta real completa, então, o conjunto de x para o qual $f(x-r) > 0$ dependerá de r. Este exemplo ilustra isto.

Definição: Seja $f(x)$ qualquer fdp. Então, para qualquer $\sigma > 0$, a família de fdps $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$, indexada pelo parâmetro σ, é chamada de família de escala com fdp padrão $f(x)$ e σ é chamada de parâmetro de escala da família.

Comentário: * O efeito de introduzir o parâmetro de escala σ é o de alongar $(\sigma > 1)$ ou contrair $(\sigma < 1)$ o gráfico de $f(x)$ enquanto ainda mantém o mesmo formato básico do gráfico.

EXEMPLO 1: Família Normal

Por exemplo, se $\mu = 0$ e definimos

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} x^2 \right\}, \quad -\infty < x < \infty,$$

então, a família de escala com fdp padrão $f(x)$ é o conjunto de distribuições normais com média $\mu = 0$ e variância σ^2 desconhecida. Basta substituir x por $\frac{x}{\sigma}$ e teremos

$$f(x|\sigma) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \frac{x^2}{\sigma^2} \right\}, \quad -\infty < x < \infty \text{ e } \sigma > 0.$$

⇒ GRÁFICO no R.

Definição: Seja $f(x)$ qualquer fdp. Então, para qualquer μ, $-\infty < \mu < \infty$, e qualquer $\sigma > 0$, a família de fdps $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, indexada pelo parâmetro (μ, σ), é chamada de família de locação - escala com fdp padrão $f(x)$; μ é chamado de parâmetro de locação e σ de parâmetro de escala.

Comentário: * O efeito de introduzir os parâmetros de locação e de escala é o de alongar $(\sigma > 1)$ ou contrair $(\sigma < 1)$ o gráfico com o parâmetro de escala e, então, modificá-lo, de modo que o ponto que estava acima de 0 fique agora acima de μ.

$\boxed{\sigma = 1}$ Família locação

$\boxed{\mu = 0}$ Família escala

EXEMPLO 1: Família Normal

Considere $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} x^2\right\}$, $-\infty < x < \infty$,

então, a família de locação - escala com fdp padrão $f(x)$ é o conjunto de distribuições normais com média μ desconhecida e variância σ^2 desconhecida. Note que $f(x \mid \mu, \sigma) =$

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty \text{ e } \sigma > 0.$$

⇒ GRÁFICO NO R.

Teorema: Seja $f(\cdot)$ qualquer fdp e seja μ um número real qualquer, e σ um número real qualquer positivo. Então X é uma variável aleatória com fdp $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, se, e ($\longleftrightarrow$) somente se, existir uma variável aleatória Z com fdp $f(z)$ e $X = \sigma Z + \mu$.

PROVA:

1- X é uma variável aleatória, tal que $X = \sigma Z + \mu$, com fdp $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) \Rightarrow$

$\exists$ uma variável aleatória Z com fdp $f(z)$.

Considere $Z = g(X) = \frac{X-\mu}{\sigma}$, então

$$P(Z \le z) = P\left(\frac{X-\mu}{\sigma} \le z\right) = P(X \le z\sigma + \mu) =$$

$$F_X(z\sigma + \mu).$$

$$f_Z(z) = f_X(z\sigma + \mu)\,\sigma = \frac{1}{\cancel{\sigma}} f\left(\frac{z\sigma + \mu - \mu}{\cancel{\sigma}}\right)\cancel{\sigma} =$$

$f(z)$ // cqd

A variável aleatória $\underset{\frac{X-\mu}{\sigma}}{\overset{Z}{\shortparallel}}$ tem fdp $f(z)$.

Isto é, a distribuição de Z é aquele membro da família locação-escala ...

correspondente a $\mu = 0$, $\sigma = 1$.

a - $\exists$ uma variável aleatória Z com fdp ⑦ ⓐ

$f(z) \Rightarrow X$ é uma variável aleatória com

fdp $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, tal que $X = \sigma Z + \mu$.

Considere $X = g(Z)$, então

$$P(X \leq x) = P(\sigma Z + \mu \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) =$$

$$F_Z\left(\frac{x-\mu}{\sigma}\right).$$

$$f_X(x) = f_Z\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} = f\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} \quad /\!/_{cqd}$$

Comentários: * O teorema relata a transformação da fdp $f(x)$ que define uma família de locação - escala para a transformação de uma variável aleatória Z com fdp $f(z)$.

* A representação em termos de Z é uma ferramenta matemática útil e pode nos ajudar a entender quando uma família de locação-escala pode ser apropriada em um contexto de modelagem.

Comentários: * Se definimos $\sigma = 1$ no teorema, temos um resultado (somente, apenas) para as famílias de locação. Se definimos $\eta = 0$ no teorema, temos um resultado para as famílias de escala.

EXEMPLO 1: Famílias de locação

Se X é uma variável aleatória com fdp $f(x-\eta)$, então X pode ser representado como $X = Z + \eta$, onde Z é uma variável aleatória com fdp $f(z)$.

→ A consideração desta representação indica quando uma família de locação pode ser um modelo apropriado para uma variável observada X.

Suponha que um experimento seja projetado para medir alguma constante física η, digamos, a temperatura de uma solução. Mas existe algum erro de medição envolvido na observação, de modo que o real valor observado X é $Z + \eta$, onde Z é o erro de medição. Note que X será maior que η se $Z > 0$ para esta observação, e será menor que τ se $z < 0$.

A distribuição do erro de medição aleatório pode ser bem conhecida a partir da experiência anterior na utilização deste dispositivo de medição para medir outras soluções.

Se esta distribuição tem fdp $f(z)$, então a fdp do valor observado X é $f(x-\tau)$.

Teorema: Seja Z uma variável aleatória com fdp $f(z)$. Suponha que $E(Z)$ e $Var(Z)$ existam. Se X é uma variável aleatória com fdp $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, então

$$E(X) = \sigma E(Z) + \mu \quad e \quad Var(X) = \sigma^2 Var(Z).$$

Em particular, se $E(Z) = 0$ e $Var(Z) = 1$, então $E(X) = \mu$ e $Var(X) = \sigma^2$.

Comentários: * Para qualquer família de locação-escala com média e variância finitas, a fdp padrão $f(z)$ pode ser escolhida de tal maneira que $E(Z) = 0$ e $Var(Z) = 1$. Isto resulta na conveniente interpretação de μ e σ^2 como a média e a variância de X, respectivamente. Este é o caso para a definição usual da família normal. (Contudo, esta não é a escolha para a definição usual da família exponencial dupla. Neste caso, $Var(Z) = 2$.)

* As probabilidades para qualquer membro de uma família de locação-escala podem ser calculadas em termos da variável padrão Z porque $P(X \leq x) = P(\sigma Z + \mu \leq x) =$

$P\left(Z \leq \frac{x-\mu}{\sigma} \right)$. Deste modo, se

$P(Z \leq z)$ for tabulado ou facilmente calculável para a variável padrão Z, então as probabilidades para X podem ser obtidas.

Os cálculos das probabilidades normais utilizando a tabela normal padrão são exemplos disto.

⟹ Comentários Importantes : Resultado útil para a construção de IC's (método da quantidade pivotal)

* Famílias de locação

Note que a distribuição de $X-\mu$ não depende de μ. $Q = \hat{\mu} - \mu$ é uma quantidade pivotal para μ.

* Famílias de escala

Note que a distribuição de X/σ não depende de σ. $Q = \hat{\sigma}/\sigma$ é uma quantidade pivotal para σ.

* Famílias de locação-escala

Note que a distribuição de $\frac{X-\mu}{\sigma}$ não depende de (μ, σ).

$$Q = \frac{\hat{\mu} - \mu}{\hat{\sigma}}$$ é uma quantidade pivotal para μ.