Probabilidad y Estadística para Inteligencia Artificial

Dr. Ing. Pablo Briff

Laboratorio de Sistemas Embebidos - FIUBA

pbriff@fi.uba.ar

8 de Agosto de 2020

Tabla de Contenidos I

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:
- Tomamos datos, a través de un muestreo repetitivo

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:
- Tomamos datos, a través de un muestreo repetitivo
- Los datos son generados aleatoriamente, ya sea a través de un proceso simulado o natural

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:
- Tomamos datos, a través de un muestreo repetitivo
- Los datos son generados aleatoriamente, ya sea a través de un proceso simulado o natural
- Hacemos hipótesis acerca del proceso que genera los datos, por ej Gaussiana i.i.d

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:
- Tomamos datos, a través de un muestreo repetitivo
- Los datos son generados aleatoriamente, ya sea a través de un proceso simulado o natural
- Hacemos hipótesis acerca del proceso que genera los datos, por ej Gaussiana i.i.d
- El proceso generador de datos está asociado con algún objeto de interés, por ej. un parámetro, o una distribución

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:
- Tomamos datos, a través de un muestreo repetitivo
- Los datos son generados aleatoriamente, ya sea a través de un proceso simulado o natural
- Hacemos hipótesis acerca del proceso que genera los datos, por ej Gaussiana i.i.d
- El proceso generador de datos está asociado con algún objeto de interés, por ej. un parámetro, o una distribución
- Este objeto es desconocido pero constante, y nos interesa encontrarlo

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Enfoque Frecuentista

- Hasta ahora hemos tomado un enfoque frecuentista de la probabilidad y estadística
- En el enfoque frecuentista, hacemos lo siguiente:
- Tomamos datos, a través de un muestreo repetitivo
- Los datos son generados aleatoriamente, ya sea a través de un proceso simulado o natural
- Hacemos hipótesis acerca del proceso que genera los datos, por ej Gaussiana i.i.d
- El proceso generador de datos está asociado con algún objeto de interés, por ej. un parámetro, o una distribución
- Este objeto es desconocido pero constante, y nos interesa encontrarlo
- Encontramos una estimación de este objeto, o hacemos un test de hipótesis con este objeto

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

- La estimación Bayesiana trata al parámetro, por ej θ, como una v.a. que toma valores Θ

- La estimación Bayesiana trata al parámetro, por ej θ, como una v.a. que toma valores Θ
- La información y creencias acerca de los valores de θ sin ninguna medición se resumen en una distribución *a priori* $\pi(\theta)$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Estimación Bayesiana

- La estimación Bayesiana trata al parámetro, por ej θ, como una v.a. que toma valores Θ
- La información y creencias acerca de los valores de θ sin ninguna medición se resumen en una distribución *a priori* $\pi(\theta)$
- Al observar datos $X = x$, se combina la información para obtener la distribución *a posteriori* $\pi(\theta|x)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Estimación Bayesiana

- La estimación Bayesiana trata al parámetro, por ej θ, como una v.a. que toma valores Θ
- La información y creencias acerca de los valores de θ sin ninguna medición se resumen en una distribución *a priori* $\pi(\theta)$
- Al observar datos $X = x$, se combina la información para obtener la distribución *a posteriori* $\pi(\theta|x)$
- Ejemplos de aplicaciones de estimación Bayesiana: filtros de spam, reconocimiento de texto, machine learning, ensayos clínicos, etc

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta)$
- $\pi(\theta|x)$ es la distribución a posteriori

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta)$
- $\pi(\theta|x)$ es la distribución a posteriori
- $f_X(x|\theta)$ es la verosimilitud

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta)$
- $\pi(\theta|x)$ es la distribución a posteriori
- $f_X(x|\theta)$ es la verosimilitud
- $\pi(\theta)$ es la distribución a priori

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta)$
- $\pi(\theta|x)$ es la distribución a posteriori
- $f_X(x|\theta)$ es la verosimilitud
- $\pi(\theta)$ es la distribución a priori
- La función $f_X(x)$ no depende de θ y es un factor de normalización

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Distribuciones A Priori y A Posteriori

- Por Bayes tenemos:
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Caso continuo: $f_X(x) = \int f_X(x|\theta)\pi(\theta)d\theta$
- Caso discreto: $\sum_i f_X(x|\theta_i)\pi(\theta_i)$
- Entonces tenemos:
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta)$
- $\pi(\theta|x)$ es la distribución a posteriori
- $f_X(x|\theta)$ es la verosimilitud
- $\pi(\theta)$ es la distribución a priori
- La función $f_X(x)$ no depende de θ y es un factor de normalización
- Los datos *entran* a través de la verosimilitud

Ejemplo

- $\pi(\theta)$ es una creencia a priori, y $\pi(\theta|x)$ es una creencia a posteriori

Ejemplo

- $\pi(\theta)$ es una creencia a priori, y $\pi(\theta|x)$ es una creencia a posteriori
- Ejemplo:

Ejemplo

- $\pi(\theta)$ es una creencia a priori, y $\pi(\theta|x)$ es una creencia a posteriori
- Ejemplo:
- Tengo 3 monedas en el bolsillo: la moneda 1 está *cargada* 3:1 a favor de ceca, la moneda 2 está balanceada, y la moneda 3 está cargada 3:1 a favor de cara

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejemplo

- $\pi(\theta)$ es una creencia a priori, y $\pi(\theta|x)$ es una creencia a posteriori
- Ejemplo:
- Tengo 3 monedas en el bolsillo: la moneda 1 está *cargada* 3:1 a favor de ceca, la moneda 2 está balanceada, y la moneda 3 está cargada 3:1 a favor de cara
- Si saco una moneda al azar y la tiro y me da cara, ¿cuál es la probabilidad de haber elegido la moneda 3?

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Las preguntas frecuentes en estimación Bayesiana son:

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Las preguntas frecuentes en estimación Bayesiana son:
- Cómo elegir la distribución a priori, y

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Las preguntas frecuentes en estimación Bayesiana son:
- Cómo elegir la distribución a priori, y
- Cómo calcular la densidad a posteriori $\pi(\theta|x)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Las preguntas frecuentes en estimación Bayesiana son:
- Cómo elegir la distribución a priori, y
- Cómo calcular la densidad a posteriori $\pi(\theta|x)$
- Conviene elegir una distribución a priori con colas poco pronunciadas para no limitar θ demasiado

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Las preguntas frecuentes en estimación Bayesiana son:
- Cómo elegir la distribución a priori, y
- Cómo calcular la densidad a posteriori $\pi(\theta|x)$
- Conviene elegir una distribución a priori con colas poco pronunciadas para no limitar θ demasiado
- Cuando no se tiene información suficiente sobre θ, usaremos distribuciones no informativas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Cómo Elegir una Distribución A Priori

- Volviendo a la expresión de Bayes
- $\pi(\theta|x) = \frac{f_X(x|\theta)\pi(\theta)}{f_X(x)}$
- Las preguntas frecuentes en estimación Bayesiana son:
- Cómo elegir la distribución a priori, y
- Cómo calcular la densidad a posteriori $\pi(\theta|x)$
- Conviene elegir una distribución a priori con colas poco pronunciadas para no limitar θ demasiado
- Cuando no se tiene información suficiente sobre θ, usaremos distribuciones no informativas
- En general, también conviene – por simplicidad – usar distribuciones conjugadas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor
- Esto es una distribución a priori no informativa

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor
- Esto es una distribución a priori no informativa
- Por ej, una distribución uniforme, $\pi(\theta) = 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor
- Esto es una distribución a priori no informativa
- Por ej, una distribución uniforme, $\pi(\theta) = 1$
- La distribución a priori puede no integrar a 1, o inclusive puede no ser integrable

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor
- Esto es una distribución a priori no informativa
- Por ej, una distribución uniforme, $\pi(\theta) = 1$
- La distribución a priori puede no integrar a 1, o inclusive puede no ser integrable
- Ejemplo: $\pi(\theta) = 1$ con $\theta \in \Re$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor
- Esto es una distribución a priori no informativa
- Por ej, una distribución uniforme, $\pi(\theta) = 1$
- La distribución a priori puede no integrar a 1, o inclusive puede no ser integrable
- Ejemplo: $\pi(\theta) = 1$ con $\theta \in \Re$
- En este caso $\int_{-\infty}^{+\infty} \pi(\theta) d\theta = \infty$

Distribución A Priori No-Informativa

- Cuando no disponemos de información acerca del parámetro θ elegimos una distribución a priori que no favorezca ningún valor
- Esto es una distribución a priori no informativa
- Por ej, una distribución uniforme, $\pi(\theta) = 1$
- La distribución a priori puede no integrar a 1, o inclusive puede no ser integrable
- Ejemplo: $\pi(\theta) = 1$ con $\theta \in \Re$
- En este caso $\int_{-\infty}^{+\infty} \pi(\theta) d\theta = \infty$
- Sin embargo, la distribución a posteriori aún puede ser que integre a 1 (y esto es lo que importa)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:
- $f_X(x, \alpha, \beta) = K\, x^{\alpha-1}(1-x)^{\beta-1}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:
- $f_X(x, \alpha, \beta) = K\, x^{\alpha-1}(1-x)^{\beta-1}$
- donde $K = 1/B(\alpha, \beta)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:
- $f_X(x, \alpha, \beta) = K\, x^{\alpha-1}(1-x)^{\beta-1}$
- donde $K = 1/B(\alpha, \beta)$
- $B(\cdot)$ es la función Beta, dada por:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:
- $f_X(x, \alpha, \beta) = K\, x^{\alpha-1}(1-x)^{\beta-1}$
- donde $K = 1/B(\alpha, \beta)$
- $B(\cdot)$ es la función Beta, dada por:
- $B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:
- $f_X(x, \alpha, \beta) = K\, x^{\alpha-1}(1-x)^{\beta-1}$
- donde $K = 1/B(\alpha, \beta)$
- $B(\cdot)$ es la función Beta, dada por:
- $B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$
- $\Gamma(n) = (n-1)!$

Distribución Beta

- La distribución Beta tiene forma de Bernoulli multivariable
- Si X es una v.a. con distribución Beta, tiene la siguiente pdf:
- $f_X(x, \alpha, \beta) = K\, x^{\alpha-1}(1-x)^{\beta-1}$
- donde $K = 1/B(\alpha, \beta)$
- $B(\cdot)$ es la función Beta, dada por:
- $B(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$
- $\Gamma(n) = (n-1)!$
- $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx \quad \Re(z) > 0$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$
- Moda$[X] = \frac{\alpha-1}{\alpha+\beta-2}$, para $\alpha, \beta > 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $\text{Moda}[X] = \frac{\alpha-1}{\alpha+\beta-2}$, para $\alpha, \beta > 1$
- $\text{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $\text{Moda}[X] = \frac{\alpha-1}{\alpha+\beta-2}$, para $\alpha, \beta > 1$
- $\text{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Con la distribución beta, se pueden emular varias distribuciones

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $\text{Moda}[X] = \frac{\alpha-1}{\alpha+\beta-2}$, para $\alpha, \beta > 1$
- $\text{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Con la distribución beta, se pueden emular varias distribuciones
- Uniforme, con $\alpha = \beta = 1$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $\text{Moda}[X] = \frac{\alpha-1}{\alpha+\beta-2}$, para $\alpha, \beta > 1$
- $\text{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Con la distribución beta, se pueden emular varias distribuciones
- Uniforme, con $\alpha = \beta = 1$
- Aproximadamente normal, con $\alpha = \beta = 5$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Distribución Beta

- Si $X \sim B(\alpha, \beta)$, entonces:
- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $\text{Moda}[X] = \frac{\alpha-1}{\alpha+\beta-2}$, para $\alpha, \beta > 1$
- $\text{var}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Con la distribución beta, se pueden emular varias distribuciones
- Uniforme, con $\alpha = \beta = 1$
- Aproximadamente normal, con $\alpha = \beta = 5$
- Exponencial, $\alpha = 1, \beta = 3$

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_{\mathrm{X}}(\mathrm{x}|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_X(x|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$
- Proponemos una distribución a priori Beta, $\pi(\theta) = B(a, b)$

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_X(x|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$
- Proponemos una distribución a priori Beta, $\pi(\theta) = B(a, b)$
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta) \propto \theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}$, con $\theta \in (0, 1)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_X(x|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$
- Proponemos una distribución a priori Beta, $\pi(\theta) = B(a, b)$
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta) \propto \theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}$, con $\theta \in (0, 1)$
- Esta es una distribución Beta, $B(\sum_i x_i + a, n - \sum_i x_i + b)$, entonces

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_X(x|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$
- Proponemos una distribución a priori Beta, $\pi(\theta) = B(a, b)$
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta) \propto \theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}$, con $\theta \in (0, 1)$
- Esta es una distribución Beta, $B(\sum_i x_i + a, n - \sum_i x_i + b)$, entonces
- $\pi(\theta|x) = \frac{\theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}}{B(\sum_i x_i + a, n - \sum_i x_i + b)}$, con $\theta \in (0, 1)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_X(x|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$
- Proponemos una distribución a priori Beta, $\pi(\theta) = B(a, b)$
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta) \propto \theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}$, con $\theta \in (0, 1)$
- Esta es una distribución Beta, $B(\sum_i x_i + a, n - \sum_i x_i + b)$, entonces
- $\pi(\theta|x) = \frac{\theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}}{B(\sum_i x_i + a, n - \sum_i x_i + b)}$, con $\theta \in (0, 1)$
- La ventaja de que sea una distribución conjugada es que la forma matemática se mantiene entre la distribución a priori y a posteriori, facilitando el análisis

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Supongamos una verosimilitud Bernoulli multivariable, es decir
- $f_X(x|\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}$
- Proponemos una distribución a priori Beta, $\pi(\theta) = B(a,b)$
- $\pi(\theta|x) \propto f_X(x|\theta)\pi(\theta) \propto \theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}$, con $\theta \in (0,1)$
- Esta es una distribución Beta, $B(\sum_i x_i + a, n - \sum_i x_i + b)$, entonces
- $\pi(\theta|x) = \frac{\theta^{\sum_i x_i + a - 1}(1-\theta)^{n-\sum_i x_i + b - 1}}{B(\sum_i x_i + a, n - \sum_i x_i + b)}$, con $\theta \in (0,1)$
- La ventaja de que sea una distribución conjugada es que la forma matemática se mantiene entre la distribución a priori y a posteriori, facilitando el análisis
- Para elegir una distribución a priori, hay que encontrar una distribución Beta que se ajuste a los datos que tenemos

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1,1)$

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1, 1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1, 1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1, 1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1, 1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$
- Moda$[X]$ cualquier valor entre 0 y 1

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1,1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$
- Moda$[X]$ cualquier valor entre 0 y 1
- $\text{var}[X] = 1/12$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1, 1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$
- Moda$[X]$ cualquier valor entre 0 y 1
- var$[X] = 1/12$
- Para la distribución a posteriori, tenemos:

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1, 1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$
- $\text{Moda}[X]$ cualquier valor entre 0 y 1
- $\text{var}[X] = 1/12$
- Para la distribución a posteriori, tenemos:
- $E[X] = \frac{\sum_i x_i + 1}{n+2}$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1,1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$
- Moda$[X]$ cualquier valor entre 0 y 1
- var$[X] = 1/12$
- Para la distribución a posteriori, tenemos:
- $E[X] = \frac{\sum_i x_i + 1}{n+2}$
- Moda$[X] = \frac{\sum_i x_i}{n}$, i.e. estimador de máxima verosimilitud

Distribución Conjugada

- Si no tenemos información previa sobre θ es razonable usar una uniforme cero-uno como distribución a priori, es decir $\pi(\theta) = B(1,1)$
- La distribución a posteriori es entonces $B(\sum_i x_i + 1, n - \sum_i x_i + 1)$
- Para la distribución a priori, tenemos:
- $E[X] = 1/2$
- Moda$[X]$ cualquier valor entre 0 y 1
- $\mathrm{var}[X] = 1/12$
- Para la distribución a posteriori, tenemos:
- $E[X] = \frac{\sum_i x_i + 1}{n+2}$
- Moda$[X] = \frac{\sum_i x_i}{n}$, i.e. estimador de máxima verosimilitud
- $\mathrm{var}[X] = \frac{(\sum_i x_i + 1)(n - \sum_i x_i + 1)}{(n+2)^2(n+3)}$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Estimador Puntual Bayesiano

- Ahora veremos el enfoque Bayesiano del problema de estimación puntual

Estimador Puntual Bayesiano

- Ahora veremos el enfoque Bayesiano del problema de estimación puntual
- Definamos una función de costo $L(\theta, a)$, la cual modela la pérdida o costo incurrido al estimar un valor a del parámetro real θ

Estimador Puntual Bayesiano

- Ahora veremos el enfoque Bayesiano del problema de estimación puntual
- Definamos una función de costo $L(\theta, a)$, la cual modela la pérdida o costo incurrido al estimar un valor a del parámetro real θ
- Generalmente nos interesa definir funciones convexas del tipo $L(\theta, a) = (\theta - a)^2$ o $L(\theta, a) = |\theta - a|$

Estimador Puntual Bayesiano

- Ahora veremos el enfoque Bayesiano del problema de estimación puntual
- Definamos una función de costo $L(\theta, a)$, la cual modela la pérdida o costo incurrido al estimar un valor a del parámetro real θ
- Generalmente nos interesa definir funciones convexas del tipo $L(\theta, a) = (\theta - a)^2$ o $L(\theta, a) = |\theta - a|$
- Definimos la pérdida a posteriori como:

Estimador Puntual Bayesiano

- Ahora veremos el enfoque Bayesiano del problema de estimación puntual
- Definamos una función de costo $L(\theta, a)$, la cual modela la pérdida o costo incurrido al estimar un valor a del parámetro real θ
- Generalmente nos interesa definir funciones convexas del tipo $L(\theta, a) = (\theta - a)^2$ o $L(\theta, a) = |\theta - a|$
- Definimos la pérdida a posteriori como:
- $h(a) = \int L(\theta, a)\pi(\theta|\mathrm{x})d\theta$

Estimador Puntual Bayesiano

- Ahora veremos el enfoque Bayesiano del problema de estimación puntual
- Definamos una función de costo $L(\theta, a)$, la cual modela la pérdida o costo incurrido al estimar un valor a del parámetro real θ
- Generalmente nos interesa definir funciones convexas del tipo $L(\theta, a) = (\theta - a)^2$ o $L(\theta, a) = |\theta - a|$
- Definimos la pérdida a posteriori como:
- $h(a) = \int L(\theta, a)\pi(\theta|\mathrm{x})d\theta$
- El estimador de Bayes es aquel que minimiza la pérdida a posteriori

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|x) d\theta$

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|x) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|\mathrm{x}) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero
- $h'(a) = 0 \iff a \int \pi(\theta|\mathrm{x}) d\theta = \int \theta \pi(\theta|\mathrm{x}) d\theta$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|\mathrm{x}) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero
- $h'(a) = 0 \iff a \int \pi(\theta|\mathrm{x}) d\theta = \int \theta \pi(\theta|\mathrm{x}) d\theta$
- Notemos que $\int \pi(\theta|\mathrm{x}) d\theta = 1$, entonces

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|\mathrm{x}) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero
- $h'(a) = 0 \iff a \int \pi(\theta|\mathrm{x}) d\theta = \int \theta\pi(\theta|\mathrm{x}) d\theta$
- Notemos que $\int \pi(\theta|\mathrm{x}) d\theta = 1$, entonces
- $a_{opt} \triangleq E[\theta|\mathrm{x}] = \int \theta\pi(\theta|\mathrm{x}) d\theta$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|\mathrm{x}) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero
- $h'(a) = 0 \iff a \int \pi(\theta|\mathrm{x}) d\theta = \int \theta \pi(\theta|\mathrm{x}) d\theta$
- Notemos que $\int \pi(\theta|\mathrm{x}) d\theta = 1$, entonces
- $a_{opt} \triangleq E[\theta|\mathrm{x}] = \int \theta \pi(\theta|\mathrm{x}) d\theta$
- $E[\theta|\mathrm{x}]$ es la *media a posteriori*, la cual minimiza la función de pérdida $h(a)$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a-\theta)^2 \pi(\theta|\text{x}) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero
- $h'(a) = 0 \iff a \int \pi(\theta|\text{x}) d\theta = \int \theta \pi(\theta|\text{x}) d\theta$
- Notemos que $\int \pi(\theta|\text{x}) d\theta = 1$, entonces
- $a_{opt} \triangleq E[\theta|\text{x}] = \int \theta \pi(\theta|\text{x}) d\theta$
- $E[\theta|\text{x}]$ es la *media a posteriori*, la cual minimiza la función de pérdida $h(a)$
- Por completitud, definimos la *varianza a posteriori* como

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Cuadrática

- Si usamos una función de pérdida cuadrática, tenemos:
- $h(a) = \int (a - \theta)^2 \pi(\theta|\mathrm{x}) d\theta$
- Para minimizar $h(a)$, derivamos con respecto al parámetro a e igualamos a cero
- $h'(a) = 0 \iff a \int \pi(\theta|\mathrm{x}) d\theta = \int \theta \pi(\theta|\mathrm{x}) d\theta$
- Notemos que $\int \pi(\theta|\mathrm{x}) d\theta = 1$, entonces
- $a_{opt} \triangleq E[\theta|\mathrm{x}] = \int \theta \pi(\theta|\mathrm{x}) d\theta$
- $E[\theta|\mathrm{x}]$ es la *media a posteriori*, la cual minimiza la función de pérdida $h(a)$
- Por completitud, definimos la *varianza a posteriori* como
- $\mathrm{var}[\theta|\mathrm{x}] = \int (\theta - E[\theta|\mathrm{x}])^2 \pi(\theta|\mathrm{x}) d\theta$

FACULTAD DE INGENIERIA Universidad de Buenos Aires

Función de Pérdida Absoluta

- Si usamos una función de pérdida de valor absoluto, tenemos:

Función de Pérdida Absoluta

- Si usamos una función de pérdida de valor absoluto, tenemos:
- $h(a) = \int |a - \theta| \pi(\theta|x) d\theta$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Absoluta

- Si usamos una función de pérdida de valor absoluto, tenemos:
- $h(a) = \int |a - \theta| \pi(\theta|x) d\theta$
- Nuevamente minimizamos $h(a)$, y llegamos a

Función de Pérdida Absoluta

- Si usamos una función de pérdida de valor absoluto, tenemos:
- $h(a) = \int |a - \theta| \pi(\theta|x) d\theta$
- Nuevamente minimizamos $h(a)$, y llegamos a
- $h'(a) = 0 \iff \int_{-\infty}^{a} \pi(\theta|x) d\theta = \int_{a}^{+\infty} \pi(\theta|x) d\theta$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Absoluta

- Si usamos una función de pérdida de valor absoluto, tenemos:
- $h(a) = \int |a - \theta| \pi(\theta|\mathrm{x}) d\theta$
- Nuevamente minimizamos $h(a)$, y llegamos a
- $h'(a) = 0 \iff \int_{-\infty}^{a} \pi(\theta|\mathrm{x}) d\theta = \int_{a}^{+\infty} \pi(\theta|\mathrm{x}) d\theta$
- La igualdad anterior se cumple con a tal que parte la distribución $\pi(\theta|\mathrm{x})$ en dos partes iguales de área $1/2$

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Función de Pérdida Absoluta

- Si usamos una función de pérdida de valor absoluto, tenemos:
- $h(a) = \int |a - \theta| \pi(\theta|\mathrm{x}) d\theta$
- Nuevamente minimizamos $h(a)$, y llegamos a
- $h'(a) = 0 \iff \int_{-\infty}^{a} \pi(\theta|\mathrm{x}) d\theta = \int_{a}^{+\infty} \pi(\theta|\mathrm{x}) d\theta$
- La igualdad anterior se cumple con a tal que parte la distribución $\pi(\theta|\mathrm{x})$ en dos partes iguales de área $1/2$
- Es decir, $a_{opt} = \mathrm{Med}[\theta|\mathrm{x}]$ es la *mediana a posteriori*

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:
- $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x|\theta)$

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:
- $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x|\theta)$
- $\hat{\theta}_{ML}$ es un estimador de un único punto

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:
- $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x|\theta)$
- $\hat{\theta}_{ML}$ es un estimador de un único punto
- Ahora buscamos el estimador de *máximo a posteriori* (MAP) como:

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:
- $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x|\theta)$
- $\hat{\theta}_{ML}$ es un estimador de un único punto
- Ahora buscamos el estimador de *máximo a posteriori* (MAP) como:
- $\hat{\theta}_{MAP} = \arg\max_{\theta} \pi(\theta|x)$

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:
- $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x|\theta)$
- $\hat{\theta}_{ML}$ es un estimador de un único punto
- Ahora buscamos el estimador de *máximo a posteriori* (MAP) como:
- $\hat{\theta}_{MAP} = \arg\max_{\theta} \pi(\theta|x)$
- A diferencia de ML, en MAP nos valemos de $\pi(\theta|x)$ para promediar las incertidumbres de θ

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Máximo a Posteriori

- Recordemos que el estimador de máxima verosimilitud (ML) se encontraba como:
- $\hat{\theta}_{ML} = \arg\max_{\theta} f_X(x|\theta)$
- $\hat{\theta}_{ML}$ es un estimador de un único punto
- Ahora buscamos el estimador de *máximo a posteriori* (MAP) como:
- $\hat{\theta}_{MAP} = \arg\max_{\theta} \pi(\theta|x)$
- A diferencia de ML, en MAP nos valemos de $\pi(\theta|x)$ para promediar las incertidumbres de θ
- El estimador MAP es la moda de la distribución a posteriori

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 1

a) Estamos interesados en estimar la tasa de mortalidad real θ en un hospital H que está probando un nuevo tratamiento. El promedio de decesos en el país es 10 %, y la tasa de mortalidad varía entre hospitales entre el 3 y 20 %. El hospital H no tiene decesos en las primeras 10 operaciones. Calcular la media a posteriori $E[\theta|\mathrm{x}]$. (Pista: usar como distribución a priori una Beta $B(3, 27)$.)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 2

a. Encontrar el estimator MAP de θ cuando la muestra x, dado θ, tiene una distribución $X \sim \mathcal{N}(\theta, \sigma_X^2)$ y la distribución a priori de θ es $\mathcal{N}(\theta_0, \sigma_\theta^2)$
b. Simular el estimador MAP para diversos valores de σ_X^2 y σ_θ^2 y sacar conclusiones en los extremos de los valores de las varianzas

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Ejercicio 3

(a) Sean $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, 1)$. Asumamos una distribución a priori de $\mu \sim \mathcal{N}(0, \tau^{-2})$, para τ conocido.

(b) Demostrar que la media a posteriori es igual a la mediana a posteriori y ambas son iguales a $\hat{\mu} = \sum_i x_i/(n + \tau^2)$.
(Pista: usar el desarrollo demostrado en
`http://www.ams.sunysb.edu/~zhu/ams570/Bayesian_Normal.pdf`)

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Bibliografía I

- E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.
- S. M. Kay, *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- "MIT, Statistics for Applications, Bayesian Statistics." https://ocw.mit.edu/courses/mathematics/18-650-statistics-for-applications-fall-2016/lecture-slides/MIT18_650F16_Bayesian_Statistics.pdf. Accessed: 2020-05-15.
- "Cambridge, Statistical Department, Bayesian Statistics." http://www.statslab.cam.ac.uk/Dept/People/djsteaching/S1B-17-06-bayesian.pdf. Accessed: 2020-05-15.

FACULTAD DE INGENIERIA
Universidad de Buenos Aires

Bibliografía II

"MIT, Brain and Cognitive Sciences, Bayesian Statistics." https://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-07-statistics-for-brain-and-cognitive-science-fall-2016/lecture-notes/MIT9_07F16_lec10.1.pdf. Accessed: 2020-05-15.