Teoría de Probabilidad

Maestría en Estadística Aplicada

Julio Hurtado, Msc.

UTB

2020

Desigualdades en Probabilidad – Parte 7

Desigualdades en Probabilidad

- **Lecturas**
- Probability Inequalities- Lin, Zhengyan, Bai, Zhidong 2011. Springer.
- Divergencia de Kullback-Leibler https://es.wikipedia.org/wiki/Divergencia_de_Kullback-Leibler.
- Información en R acerca de la distancia de Kullback-Leibler http://www.inside-r.org/packages/cran/FNN/docs/KL.divergence
-

Desigualdades en Probabilidad

- **Lecturas**
- Probability Inequalities- Lin, Zhengyan, Bai, Zhidong 2011. Springer.
- Divergencia de Kullback-Leibler https://es.wikipedia.org/wiki/Divergencia_de_Kullback-Leibler.
- Información en R acerca de la distancia de Kullback-Leibler http://www.inside-r.org/packages/cran/FNN/docs/KL.divergence
-

Desigualdades en Probabilidad

- **Lecturas**
- `Probability Inequalities`- Lin, Zhengyan, Bai, Zhidong 2011. Springer.
- `Divergencia de Kullback-Leibler` `https://es.wikipedia.org/wiki/Divergencia_de_Kullback-Leibler`.
- `Información en R acerca de la distancia de Kullback-Leibler` `http://www.inside-r.org/packages/cran/FNN/docs/KL.divergence`
-

Desigualdades en Probabilidad

- **Lecturas**
- Probability Inequalities- Lin, Zhengyan, Bai, Zhidong 2011. Springer.
- Divergencia de Kullback-Leibler https://es.wikipedia.org/wiki/Divergencia_de_Kullback-Leibler.
- Información en R acerca de la distancia de Kullback-Leibler http://www.inside-r.org/packages/cran/FNN/docs/KL.divergence

Desigualdades en Probabilidad

- **Lecturas**
- Probability Inequalities- Lin, Zhengyan, Bai, Zhidong 2011. Springer.
- Divergencia de Kullback-Leibler https://es.wikipedia.org/wiki/Divergencia_de_Kullback-Leibler.
- Información en R acerca de la distancia de Kullback-Leibler http://www.inside-r.org/packages/cran/FNN/docs/KL.divergence
-

Desigualdades en Probabilidad

- Las desigualdades en Probabilidad, es un capítulo de la Teoría de las Probabilidades muy importante, lleno de aplicaciones teóricas y prácticas.
- Por ejemplo la *desigualdad de Hoeffding* puede aplicarse al caso especial de variables de Bernoulli idénticamente distribuidas, y esta es la manera en que se aplica frecuentemente en combinatoria y en informática, la desigualdad de Chebyshev, ofrece una cota inferior a la probabilidad de que el valor de una variable aleatoria esté a cierta distancia de sus esperanza, bajo ciertas condiciones, etc.
- En esta sección presentamos algunas de las más importantes desigualdades de la Teoría de las Probabilidades y ciertas aplicaciones que conllevan.

Desigualdades en Probabilidad

- Las desigualdades en Probabilidad, es un capítulo de la Teoría de las Probabilidades muy importante, lleno de aplicaciones teóricas y prácticas.
- Por ejemplo la *desigualdad de Hoeffding* puede aplicarse al caso especial de variables de Bernoulli idénticamente distribuidas, y esta es la manera en que se aplica frecuentemente en combinatoria y en informática, la desigualdad de Chebyshev, ofrece una cota inferior a la probabilidad de que el valor de una variable aleatoria esté a cierta distancia de sus esperanza, bajo ciertas condiciones, etc.
- En esta sección presentamos algunas de las más importantes desigualdades de la Teoría de las Probabilidades y ciertas aplicaciones que conllevan.

Desigualdades en Probabilidad

- Las desigualdades en Probabilidad, es un capítulo de la Teoría de las Probabilidades muy importante, lleno de aplicaciones teóricas y prácticas.
- Por ejemplo la *desigualdad de Hoeffding* puede aplicarse al caso especial de variables de Bernoulli idénticamente distribuidas, y esta es la manera en que se aplica frecuentemente en combinatoria y en informática, la desigualdad de Chebyshev, ofrece una cota inferior a la probabilidad de que el valor de una variable aleatoria esté a cierta distancia de sus esperanza, bajo ciertas condiciones, etc.
- En esta sección presentamos algunas de las más importantes desigualdades de la Teoría de las Probabilidades y ciertas aplicaciones que conllevan.

La desigualdad Gaussiana

- **La desigualdad Gaussiana** Sea $X \sim N(0,1)$ entonces

$$\mathbb{P}(|X| > \epsilon) \leq \frac{2e^{-\epsilon^2/2}}{\epsilon}$$

Si $X_1, \dots, X_n \sim N(0,1)$ entonces

$$\mathbb{P}(|\overline{X_n}| > \epsilon) \leq \frac{2}{\sqrt{n}\epsilon} e^{-n\epsilon^2/2}$$

-

La desigualdad Gaussiana

- **La desigualdad Gaussiana** Sea $X \sim N(0,1)$ entonces

$$\mathbb{P}(|X| > \epsilon) \leq \frac{2e^{-\epsilon^2/2}}{\epsilon}$$

Si $X_1, \ldots, X_n \sim N(0,1)$ entonces

$$\mathbb{P}(|\overline{X_n}| > \epsilon) \leq \frac{2}{\sqrt{n}\epsilon} e^{-n\epsilon^2/2}$$

-

Desigualdad de Markov

- **Desigualdad de Markov** Sea X es una variable no negativa y supongamos que $\mathbb{E}(X)$ existe.

  $$\mathbb{P}(X > t) \leq \frac{\mathbb{E}(X)}{t}.$$

  Para algún $t > 0$.

-

Desigualdad de Markov

- **Desigualdad de Markov** Sea X es una variable no negativa y supongamos que $\mathbb{E}(X)$ existe.

$$\mathbb{P}(X > t) \leq \frac{\mathbb{E}(X)}{t}.$$

  Para algún $t > 0$.

-

Desigualdad de Chebyshev

- **Desigualdad de Chebyshev** Sea $\mu = \mathbb{E}(X)$ y $\sigma^2 = V(X)$. Entonces

$$\mathbb{P}(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2} \qquad y \qquad \mathbb{P}(|Z| \geq k) \leq \frac{1}{k^2}.$$

donde $Z = (X-\mu)/\sigma$. En particular $\mathbb{P}(|Z| > 2) \leq 1/4$ y $\mathbb{P}(|Z| > 3) \leq 1/9$.

-

Desigualdad de Chebyshev

- **Desigualdad de Chebyshev** Sea $\mu = \mathbb{E}(X)$ y $\sigma^2 = V(X)$. Entonces

$$\mathbb{P}(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2} \qquad y \qquad \mathbb{P}(|Z| \geq k) \leq \frac{1}{k^2}.$$

  donde $Z = (X-\mu)/\sigma$. En particular $\mathbb{P}(|Z| > 2) \leq 1/4$ y $\mathbb{P}(|Z| > 3) \leq 1/9$.

-

Desigualdad de Hoeffding

- **Desigualdad de Hoeffding** Para hacer una demostración de este resultado se debe recordar que se dice que una función g es convexa, si para cada x, y para cada $\alpha \in [0,1]$.

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

- **Propiedad** Supongamos que $\mathbb{E}(X) = 0$ y que $a \leq x \leq b$. Entonces

$$\mathbb{E}(e^{tX}) \leq e^{t^2(b-a)^2/8}.$$

Método de Chernoff Sea X una variable aleatoria. Entonces

$$\mathbb{P}(X > \epsilon) \leq \inf_{t \geq 0} e^{-t\epsilon}\mathbb{E}(e^{tX}).$$

-

Desigualdad de Hoeffding

- **Desigualdad de Hoeffding**Para hacer una demostración de este resultado se debe recordar que se dice que una función g es convexa, si para cada x, y para cada $\alpha \in [0,1]$.

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

- **Propiedad** Supongamos que $\mathbb{E}(X) = 0$ y que $a \leq x \leq b$. Entonces

$$\mathbb{E}(e^{tX}) \leq e^{t^2(b-a)^2/8}.$$

Método de Chernoff Sea X una variable aleatoria. Entonces

$$\mathbb{P}(X > \epsilon) \leq \inf_{t \geq 0} e^{-t\epsilon}\mathbb{E}(e^{tX}).$$

Desigualdad de Hoeffding

- **Desigualdad de Hoeffding**Para hacer una demostración de este resultado se debe recordar que se dice que una función g es convexa, si para cada x, y para cada $\alpha \in [0, 1]$.

$$g(\alpha x + (1-\alpha)y) \leq \alpha g(x) + (1-\alpha)g(y).$$

- **Propiedad** Supongamos que $\mathbb{E}(X) = 0$ y que $a \leq x \leq b$. Entonces

$$\mathbb{E}(e^{tX}) \leq e^{t^2(b-a)^2/8}.$$

Método de Chernoff Sea X una variable aleatoria. Entonces

$$\mathbb{P}(X > \epsilon) \leq \inf_{t \geq 0} e^{-t\epsilon}\mathbb{E}(e^{tX}).$$

-

Desigualdad de Hoeffding

- Sean $Y_1, Y_2, \ldots, Y_n$ observaciones indepedientes, tal que $\mathbb{E}(Y_i) = 0$ y $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para algún $t > 0$
- $$\mathbb{P}(|\overline{Y_n} - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$
- La desigualdad de Hoeffding proporciona una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado.
- **Corolario** Si $X_1, X_2, \cdots, X_n$ son independientes con $\mathbb{P}(a \leq X_i \leq b) = 1$ y una media común μ, entonces con probabilidad de al menos $1 - \delta$

$$|\overline{X_n} - \mu| \leq \sqrt{\frac{(b-a)^2}{2n} \log\left(\frac{2}{\delta}\right)}.$$

Desigualdad de Hoeffding

- Sean $Y_1, Y_2, \ldots, Y_n$ observaciones indepedientes, tal que $\mathbb{E}(Y_i) = 0$ y $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para algún $t > 0$
- $$\mathbb{P}(|\overline{Y_n} - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$
- La desigualdad de Hoeffding proporciona una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado.
- **Corolario** Si $X_1, X_2, \cdots, X_n$ son independientes con $\mathbb{P}(a \leq X_i \leq b) = 1$ y una media común μ, entonces con probabilidad de al menos $1 - \delta$

$$|\overline{X_n} - \mu| \leq \sqrt{\frac{(b-a)^2}{2n} \log\left(\frac{2}{\delta}\right)}.$$

Desigualdad de Hoeffding

- Sean $Y_1, Y_2, \ldots, Y_n$ observaciones indepedientes, tal que $\mathbb{E}(Y_i) = 0$ y $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para algún $t > 0$
-
$$\mathbb{P}(|\overline{Y_n} - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$
- La desigualdad de Hoeffding proporciona una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado.
- **Corolario** Si $X_1, X_2, \cdots, X_n$ son independientes con $\mathbb{P}(a \leq X_i \leq b) = 1$ y una media común μ, entonces con probabilidad de al menos $1 - \delta$
$$|\overline{X_n} - \mu| \leq \sqrt{\frac{(b-a)^2}{2n} \log\left(\frac{2}{\delta}\right)}.$$
-

Desigualdad de Hoeffding

- Sean $Y_1, Y_2, \ldots, Y_n$ observaciones indepedientes, tal que $\mathbb{E}(Y_i) = 0$ y $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para algún $t > 0$
- $$\mathbb{P}(|\overline{Y_n} - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$
- La desigualdad de Hoeffding proporciona una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado.
- **Corolario** Si $X_1, X_2, \cdots, X_n$ son independientes con $\mathbb{P}(a \leq X_i \leq b) = 1$ y una media común μ, entonces con probabilidad de al menos $1 - \delta$
$$|\overline{X_n} - \mu| \leq \sqrt{\frac{(b-a)^2}{2n} \log\left(\frac{2}{\delta}\right)}.$$

Desigualdad de Hoeffding

- Sean $Y_1, Y_2, \ldots, Y_n$ observaciones indepedientes, tal que $\mathbb{E}(Y_i) = 0$ y $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para algún $t > 0$
- $$\mathbb{P}(|\overline{Y_n} - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$
- La desigualdad de Hoeffding proporciona una cota superior a la probabilidad de que la suma de variables aleatorias se desvíe una cierta cantidad de su valor esperado.
- **Corolario** Si $X_1, X_2, \cdots, X_n$ son independientes con $\mathbb{P}(a \leq X_i \leq b) = 1$ y una media común μ, entonces con probabilidad de al menos $1 - \delta$

$$|\overline{X_n} - \mu| \leq \sqrt{\frac{(b-a)^2}{2n} \log\left(\frac{2}{\delta}\right)}.$$

Teorema de McDiarmid

- **Ejemplo** Sea $X_1, X_2, \ldots, X_n \sim$ Bernoulli(p). Desde la desigualdad Hoeffding se tiene

$$\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq 2\epsilon^{-2n\epsilon^2}.$$

- **Teorema de McDiarmid** Sea $X_1, X_2, \ldots X_n$ variables aleatorias independientes. Suponganse que

$$\sup_{x_1,\ldots,x_n,x_i^{'}} \Big| g(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n) - g(x_1,\ldots,x_{i-1},x_i^{'},x_{i+1},\ldots,x_n)$$

para $i = 1, \ldots, n$. Entonces

$$\mathbb{P}\Big(g(X_1,\ldots,X_n) - \mathbb{E}(g(X_1,\ldots,X_n)) \geq \epsilon\Big) \leq \exp\Big\{-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}\Big\}.$$

-

Teorema de McDiarmid

- **Ejemplo** Sea $X_1, X_2, \dots, X_n \sim$ Bernoulli(p). Desde la desigualdad Hoeffding se tiene

$$\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq 2\epsilon^{-2n\epsilon^2}.$$

- **Teorema de McDiarmid** Sea $X_1, X_2, \dots X_n$ variables aleatorias independientes. Suponganse que

$$\sup_{x_1,\dots,x_n,x_i^{'}} \left| g(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n) - g(x_1,\dots,x_{i-1},x_i^{'},x_{i+1},\dots,x_n) \right.$$

para $i = 1, \dots, n$. Entonces

$$\mathbb{P}\Big(g(X_1,\dots,X_n) - \mathbb{E}(g(X_1,\dots,X_n)) \geq \epsilon\Big) \leq \exp\left\{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}\right\}.$$

Teorema de McDiarmid

- **Ejemplo** Sea $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$. Desde la desigualdad Hoeffding se tiene

$$\mathbb{P}(|\overline{X_n} - p| > \epsilon) \leq 2\epsilon^{-2n\epsilon^2}.$$

- **Teorema de McDiarmid** Sea $X_1, X_2, \ldots X_n$ variables aleatorias independientes. Suponganse que

$$\sup_{x_1,\ldots,x_n,x_i^{'}} \Big| g(x_1,\ldots,x_{i-1},x_i,x_{i+1},\ldots,x_n) - g(x_1,\ldots,x_{i-1},x_i^{'},x_{i+1},\ldots,x_n)$$

para $i = 1, \ldots, n$. Entonces

$$\mathbb{P}\Big(g(X_1,\ldots,X_n) - \mathbb{E}(g(X_1,\ldots,X_n)) \geq \epsilon\Big) \leq \exp\Big\{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}\Big\}.$$

-

Teorema de McDiarmid

- **Ejemplo** Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.
- Sea M el número de recipientes vacios y sea $F = M/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces
-
$$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1-1/n)^m = ne^{m\log(1-1/n)} \approx ne^{-m/n}.$$
- y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?.
- Para ello definamos las variables $X_1, \ldots, X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1, \ldots, X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ muestra que

$$\mathbb{P}(|Z-\mu| > t) \leq 2e^{-2t^2/m}$$

Teorema de McDiarmid

- **Ejemplo** Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.
- Sea M el número de recipientes vacios y sea $F = M/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces
- $$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1-1/n)^m = ne^{m\log(1-1/n)} \approx ne^{-m/n}.$$
- y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?.
- Para ello definamos las variables $X_1, \ldots, X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1, \ldots, X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ muestra que

$$\mathbb{P}(|Z-\mu| > t) \leq 2e^{-2t^2/m}$$

Teorema de McDiarmid

- **Ejemplo** Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.
- Sea M el número de recipientes vacios y sea $F = M/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces
-
$$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1 - 1/n)^m = ne^{m \log(1-1/n)} \approx ne^{-m/n}.$$
- y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?.
- Para ello definamos las variables $X_1, \ldots, X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1, \ldots, X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ muestra que

$$\mathbb{P}(|Z - \mu| > t) \leq 2e^{-2t^2/m}$$

Teorema de McDiarmid

- **Ejemplo** Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.
- Sea M el número de recipientes vacios y sea $F = M/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces
- $$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1-1/n)^m = ne^{m\log(1-1/n)} \approx ne^{-m/n}.$$
- y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?.
- Para ello definamos las variables $X_1, \ldots, X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1, \ldots, X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ muestra que

$$\mathbb{P}(|Z-\mu| > t) \leq 2e^{-2t^2/m}$$

Teorema de McDiarmid

- **Ejemplo** Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.
- Sea M el número de recipientes vacios y sea $F = M/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces
- $$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1 - 1/n)^m = ne^{m \log(1-1/n)} \approx ne^{-m/n}.$$
- y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?.
- Para ello definamos las variables $X_1, \ldots, X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1, \ldots, X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ muestra que

$$\mathbb{P}(|Z - \mu| > t) \leq 2e^{-2t^2/m}.$$

Teorema de McDiarmid

- **Ejemplo** Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.
- Sea M el número de recipientes vacios y sea $F = M/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces
- $$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1-1/n)^m = ne^{m\log(1-1/n)} \approx ne^{-m/n}.$$
- y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?.
- Para ello definamos las variables $X_1, \dots, X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1, \dots, X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ muestra que

$$\mathbb{P}(|Z-\mu| > t) \leq 2e^{-2t^2/m}.$$

Desigualdad de Cauchy -Schwartz

- Como las fracción de recipientes vacios es $F = Z/m$ con media $\theta = \mu/n$. Tenemos
- $$\mathbb{P}(|F-\theta| > t) = \mathbb{P}(|Z-\mu| > nt) \leq 2e^{-2n^2t^2/m}.$$
- **Desigualdad de Cauchy -Schwartz** Si X y Y tienen varianza finita, entonces
$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$
- La desigualdad de Cauchy-Schwarz puede ser escrita como
$$\mathrm{Cov}^2(X,Y) \leq \delta_X^2 \delta_Y^2.$$
-

Desigualdad de Cauchy -Schwartz

- Como las fracción de recipientes vacios es $F = Z/m$ con media $\theta = \mu/n$. Tenemos
- $$\mathbb{P}(|F-\theta| > t) = \mathbb{P}(|Z-\mu| > nt) \leq 2e^{-2n^2t^2/m}.$$
- **Desigualdad de Cauchy -Schwartz** Si X y Y tienen varianza finita, entonces
$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$
- La desigualdad de Cauchy-Schwarz puede ser escrita como
$$\mathrm{Cov}^2(X,Y) \leq \delta_X^2\delta_Y^2.$$
-

Desigualdad de Cauchy -Schwartz

- Como las fracción de recipientes vacios es $F = Z/m$ con media $\theta = \mu/n$. Tenemos
-
$$\mathbb{P}(|F-\theta|>t)=\mathbb{P}(|Z-\mu|>nt)\leq 2e^{-2n^2t^2/m}.$$
- **Desigualdad de Cauchy -Schwartz** Si X y Y tienen varianza finita, entonces
$$\mathbb{E}|XY|\leq\sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$
- La desigualdad de Cauchy-Schwarz puede ser escrita como
$$\mathrm{Cov}^2(X,Y)\leq\delta_X^2\delta_Y^2.$$
-

Desigualdad de Cauchy -Schwartz

- Como las fracción de recipientes vacios es $F = Z/m$ con media $\theta = \mu/n$. Tenemos
- $$\mathbb{P}(|F - \theta| > t) = \mathbb{P}(|Z - \mu| > nt) \leq 2e^{-2n^2t^2/m}.$$
- **Desigualdad de Cauchy -Schwartz** Si X y Y tienen varianza finita, entonces
$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$
- La desigualdad de Cauchy-Schwarz puede ser escrita como
$$\mathrm{Cov}^2(X, Y) \leq \delta_X^2 \delta_Y^2.$$

Desigualdad de Cauchy -Schwartz

- Como las fracción de recipientes vacios es $F = Z/m$ con media $\theta = \mu/n$. Tenemos
- $$\mathbb{P}(|F-\theta| > t) = \mathbb{P}(|Z-\mu| > nt) \leq 2e^{-2n^2t^2/m}.$$
- **Desigualdad de Cauchy -Schwartz** Si X y Y tienen varianza finita, entonces
$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$
- La desigualdad de Cauchy-Schwarz puede ser escrita como
$$\mathrm{Cov}^2(X,Y) \leq \delta_X^2\delta_Y^2.$$
-

Desigualdad de Jensen

- **Desigualdad de Jensen** Si g es convexa, entonces

$$\mathbb{E}g(X) \geq g(\mathbb{E}X)$$

- Si g es cóncava

$$\mathbb{E}g(X) \leq g(\mathbb{E}X).$$

- **Ejemplo** La distancia de Kullbacker-Leibler entre una distribución Gaussiana p con media μ_1 y varianza σ_1^2 y una distribución Gaussiana q con media μ_2 y varianza σ_2^2 es dado por

$$D(p,q) = \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

-

Desigualdad de Jensen

- **Desigualdad de Jensen** Si g es convexa, entonces

$$\mathbb{E}g(X) \geq g(\mathbb{E}X)$$

- Si g es cóncava

$$\mathbb{E}g(X) \leq g(\mathbb{E}X).$$

- **Ejemplo** La distancia de Kullbacker-Leibler entre una distribución Gaussiana p con media μ_1 y varianza σ_1^2 y una distribución Gaussiana q con media μ_2 y varianza σ_2^2 es dado por

$$D(p,q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

Desigualdad de Jensen

- **Desigualdad de Jensen** Si g es convexa, entonces

$$\mathbb{E}g(X) \geq g(\mathbb{E}X)$$

- Si g es cóncava

$$\mathbb{E}g(X) \leq g(\mathbb{E}X).$$

- **Ejemplo** La distancia de Kullbacker-Leibler entre una distribución Gaussiana p con media μ_1 y varianza σ_1^2 y una distribución Gaussiana q con media μ_2 y varianza σ_2^2 es dado por

$$D(p, q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

Desigualdad de Jensen

- **Desigualdad de Jensen** Si g es convexa, entonces

$$\mathbb{E}g(X) \geq g(\mathbb{E}X)$$

- Si g es cóncava

$$\mathbb{E}g(X) \leq g(\mathbb{E}X).$$

- **Ejemplo** La distancia de Kullbacker-Leibler entre una distribución Gaussiana p con media μ_1 y varianza σ_1^2 y una distribución Gaussiana q con media μ_2 y varianza σ_2^2 es dado por

$$D(p, q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}.$$

Desigualdad de Jensen

- **Teorema** Supongamos que $X_n \geq 0$ y para cada $\epsilon > 0$,

$$\mathbb{P}(X_n > \epsilon) \leq c_1 e^{-c_2 n \epsilon^2}$$

para algún $c_2 > 0$ y $c_1 > 1/e$. Entonces

$$\mathbb{E}(X_n) \leq \sqrt{\frac{C}{n}}.$$

donde $C = (1 + \log(c_1))/c_2$

- **Teorema** Sean $X_1, \ldots, X_n$ variables aleatorias. Supongamos que existe un $\delta > 0$ tal que $\mathbb{E}(e^{tX_i}) \leq e^{t^2\delta^2/2}$ para todo $t > 0$. Entonces

$$\mathbb{E}\left(\max_{1 \leq i \leq n} X_i\right) \leq \delta\sqrt{2 \log n}.$$

-

Desigualdad de Jensen

- **Teorema** Supongamos que $X_n \geq 0$ y para cada $\epsilon > 0$,

$$\mathbb{P}(X_n > \epsilon) \leq c_1 e^{-c_2 n \epsilon^2}$$

para algún $c_2 > 0$ y $c_1 > 1/e$. Entonces

$$\mathbb{E}(X_n) \leq \sqrt{\frac{C}{n}}.$$

donde $C = (1 + \log(c_1))/c_2$

- **Teorema** Sean $X_1, \ldots, X_n$ variables aleatorias. Supongamos que existe un $\delta > 0$ tal que $\mathbb{E}(e^{tX_i}) \leq e^{t^2\delta^2/2}$ para todo $t > 0$. Entonces

$$\mathbb{E}\left(\max_{1 \leq i \leq n} X_i\right) \leq \delta\sqrt{2 \log n}.$$

Desigualdad de Jensen

- **Teorema** Supongamos que $X_n \geq 0$ y para cada $\epsilon > 0$,

$$\mathbb{P}(X_n > \epsilon) \leq c_1 e^{-c_2 n \epsilon^2}$$

para algún $c_2 > 0$ y $c_1 > 1/e$. Entonces

$$\mathbb{E}(X_n) \leq \sqrt{\frac{C}{n}}.$$

donde $C = (1 + \log(c_1))/c_2$

- **Teorema** Sean $X_1, \ldots, X_n$ variables aleatorias. Supongamos que existe un $\delta > 0$ tal que $\mathbb{E}(e^{tX_i}) \leq e^{t^2\delta^2/2}$ para todo $t > 0$. Entonces

$$\mathbb{E}\left(\max_{1 \leq i \leq n} X_i\right) \leq \delta\sqrt{2 \log n}.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que
  $$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que

$$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que
  $$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que

$$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que

$$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que

$$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- En Estadística, Probabilidades y Machine Learning, usamos la notación o_P y O_P. Por ejemplo, $a_n = o(1)$ significa que $a_n \to 0$ cuando $n \to \infty$. $a_n = o(b_n)$ significa $a_n/b_n = o(1)$.
- $a_n = O(1)$ (eventualmente acotados) significa que para un n una cantidad muy grande, $|a_n| \leq C$ para algún $C > 0$. $a_n = O(b_n)$, significa que $a_n/b_n = O(1)$.
- Escribimos $a_n \sim b_n$ si ambos a/b y b/a son eventualmente acotados ($a_n = \Theta(b_n)$). Veamos estas cosas en el terreno de las probabilidades, es decir $Y_n = o_P(1)$ significa que para cada $\epsilon > 0$,
- $$\mathbb{P}(|Y_n| > \epsilon) \to 0.$$
- Decimos que $Y_n = o_P(a_n)$ si, $Y_n/a_n = o_P(1)$.
- Si tenemos $Y_n = O_P(1)$ si, para cada $\epsilon > 0$, existe un $C > 0$ tal que

$$\mathbb{P}(|Y_n| > C) \leq \epsilon.$$

Sucesiones acotadas

- Se dice que $Y_n = O_P(a_n)$ si $Y_n/a_n = O_p(1)$. Sean $Y_1, Y_2, \ldots, Y_n$ lanzamientos de monedas, esto es $Y_i \in \{0,1\}$. Sea $p = \mathbb{P}(Y_i = 1)$. Sea

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

  Se prueba que: $\hat{p}_n - p = o_P(1)$ y $\hat{p}_n - p = O_P(1/\sqrt{n})$.

-

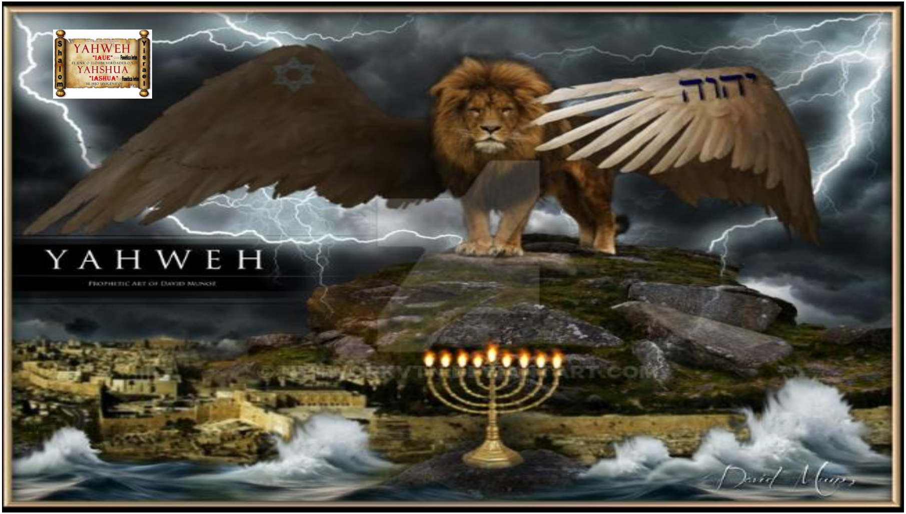
Sucesiones acotadas

- Se dice que $Y_n = O_P(a_n)$ si $Y_n/a_n = O_p(1)$. Sean $Y_1, Y_2, \ldots, Y_n$ lanzamientos de monedas, esto es $Y_i \in \{0,1\}$. Sea $p = \mathbb{P}(Y_i = 1)$. Sea

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Se prueba que: $\hat{p}_n - p = o_P(1)$ y $\hat{p}_n - p = O_P(1/\sqrt{n})$.

-


YAHWEH
"IAUE"
YAHSHUA
"IASHUA"
Shalom
Yisrael
יהוה
YAHWEH
PROPHETIC ART OF DAVID MUNOZ