Probability Theory Lecture 10

Padmini Mukkamala

Budapest University of Technology and Economics

November 12, 2021

Overview

1. Probability Bounds
2. Optimizations
3. Limits and Limit Distributions

Probability Bounds - 1

Markov Inequality

Given a **positive** random variable X ($P(X < 0) = 0$), the following is true for any real number $a > 0$:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Proof: $\int_0^\infty x f_X(x) = E(X) \implies \int_0^a x f_X(x) + \int_a^\infty x f_X(x) = E(X)$
$\implies \int_a^\infty x f_X(x) \leq E(X)$
$\implies a \int_a^\infty f_X(x) \leq E(X)$
$\implies aP(X \geq a) \leq E(X)$

Example

Let $X \sim Pois(5)$. We will estimate $P(X \geq 10)$. (Note that we can find the exact probability).

Example

Let $X \sim Pois(5)$. We will estimate $P(X \geq 10)$. (Note that we can find the exact probability).
Using Markov's inequality, we get an upper bound of 0.5, while the actual value is 0.0318.

Probability Bounds - 2

Chebyshev's Inequality

Given **any** random variable X with mean μ and standard deviation σ, the following is true for any real number $a > 0$:

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Another way of writing it using the **standardization** of X:

$$P(\left|\frac{X - \mu}{\sigma}\right| \geq a) \leq \frac{1}{a^2}$$

In other words, the probability of X being within 2, 3, 4 standard deviations of the mean are $75\%, 89\%$ and 93.75%.

Probability Bounds - 2 cont.

Proof: We notice that $(X-\mu)^2$ is a positive random variable and we can apply Markov's inequality to it.

Example

Let $X \sim Pois(5)$. We will estimate $P(X \geq 10)$. (Note that we can find the exact probability).

Example

Let $X \sim Pois(5)$. We will estimate $P(X \geq 10)$. (Note that we can find the exact probability).
Using Chebyshev's inequality, we get an upper bound of 0.2, while the actual value is 0.0318.

Probability Bounds - 3

Chernoff's bounds

Given **any** random variable X, the following is true for any real number a (not necessarily positive):

$$P(X \geq a) \leq \frac{M_X(t)}{e^{ta}}, \forall t > 0$$

We can optimize t to get the best bound.

Proof: $P(X \geq a) = P(Xt \geq ta) = P(e^{Xt} \geq e^{at})$.
We note here that since e^{Xt} is a positive random variable, we can again apply Markov's Inequality.
$P(e^{Xt} \geq e^{at}) \leq \frac{E(e^{Xt})}{e^{at}}$, where we note that the numerator is the MGF.

Example

Let $X \sim Pois(5)$. We will estimate $P(X \geq 10)$. (Note that we can find the exact probability).

Example

Let $X \sim Pois(5)$. We will estimate $P(X \geq 10)$. (Note that we can find the exact probability).
Using Chernoff's bounds, we get an upper bound of 0.1449, while the actual value is 0.0318.

Distance from mean - 1

Median

For any random variable X, $E(|X - c|)$ is minimized when c is the median, that is, $\int_c^\infty f_X(x)dx = \frac{1}{2}$.

Proof: $\int_{-\infty}^\infty |x - c| f_X(x)dx$
$= \int_{-\infty}^c (c - x) f_X(x)dx + \int_c^\infty (x - c) f_X(x)dx$
We will use the fact that $\frac{d}{dc}\int_{-\infty}^c g(x)dx = g(c)$ to differentiate the above equation with respect to c. Then we will get,
$\int_{-\infty}^c f_X(x)dx + cf_X(c) - cf_X(c) - cf_X(c) - \int_c^\infty f_X(x)dx + cf_X(c) = 0$
which gives us, $\int_{-\infty}^c f_X(x)dx = \int_c^\infty f_X(x)dx$ which is true for the median of the random variable.

Distance from mean - 2

Steiner Equality

For any random variable X, $E((X-c)^2)$ is minimized when $c=\mu$, where $\mu = E(X)$.

Proof:

$$E((X-c)^2) = E(((X-\mu)+(\mu-c))^2) = Var(X) + (\mu-c)^2.$$

Independent Identical Distributions

Definition

A sequence of random variables $X_1, X_2, X_3, ..., X_n$ is said to be **i.i.d**, or "Independent Identically Distributed" if they all have the same probability distribution and if they are totally independent. (In some cases pairwise independence is sufficient).

Example: $Bin(n, p)$ is the sum of n Bernoulli $1(p)$ i.i.d. random variables.

Expected Value and Standard Deviation

If $X_1, X_2, ..., X_n$ are i.i.d. random variables, each with mean μ and standard deviation σ, then their sum $X = \sum_{i=1}^{n} X_i$ is a random variable with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$.

Weak Law of Large Numbers

Weak Law of Large Numbers

Given a sequence $X_1, X_2, ...,$ of i.i.d random variables (pairwise independence is sufficient) with mean μ and standard deviation σ, we define a new sequence of averages, $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$. Then the following is true for all $\epsilon > 0$:

$$\lim_{n\to\infty} P(|\overline{X_n} - \mu| > \epsilon) = 0$$

Proof: We notice that the mean of $\overline{X_n}$ is μ and its standard deviation is $\frac{\sigma}{\sqrt{n}}$. So, using Chebyshev's,
$P(|\overline{X_n} - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$
and we can see that this tends to 0 as $n \to \infty$.

Weak Law of Large Numbers

Let A be an event for a certain experiment. We repeat the experiment, each occurance of the experiment being independent and identical. Let 1_A^i denote the indicator random variable for the event A in the i^{th} occurance of the experiment, i.e,

$$1_A^i = \begin{cases} 1 & \text{if A occured} \\ 0 & \text{if A did not occur} \end{cases}.$$

We know that $E(1_A^i) = P(A)$.

If we look at $X_n = \frac{\sum_{i=0}^n 1_A^i}{n}$, then this precisely denotes the relative frequency of the event A in n independent occurances of the experiment.

Then the Weak Law states that as $n \to \infty$, the relative frequency tends to the probability of the event!

Strong Law of Large Numbers

Strong Law of Large Numbers

Given a sequence $X_1, X_2, ...,$ of i.i.d random variables with mean μ, we define a new sequence of averages, $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$. Then the following is true for all $\epsilon > 0$:

$$P(\lim_{n\to\infty} \overline{X_n} \to \mu) = 1$$

or in other words, $\overline{X_n}$ almost surely (with probability 1) converges to the mean μ.

Central Limit Theorem

Central Limit Theorem

Given a sequence $X_1, X_2, ...,$ of i.i.d random variables with mean μ and standard deviation σ, we define a new sequence of **standardized random variables**, $Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$. Then the limiting random variable $Z = \lim_{n\to\infty} Z_n$ has the standard normal distribution.

$$Z_n \to N(0,1)$$

For a $n > 20$ it is standard practice to approximate the distribution of Z_n with the standard normal distribution $N(0,1)$.

Central Limit Theorem

Proof

Since, $Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$, this can also be written as,
$Z_n = \frac{\sum_{i=1}^n \frac{X_i-\mu}{\sigma}}{\sqrt{n}}$.
Lets define new variables $X_i' = \frac{X_i-\mu}{\sigma}$. Then $Z_n = \frac{\sum X_i'}{\sqrt{n}}$.
Let the moment generating function of X_i' be $M(t)$. Then since X_i''s are standardized, $M(0) = 1$, $M'(0) = 0$, $M''(0) = 1$.
Then, $M_{Z_n}(t) = (M(\frac{t}{\sqrt{n}}))^n$.
What we need to show is that $\lim_{n\to\infty} Z_n \to N(0,1)$. Since the MGF of $N(0,1)$ is $e^{\frac{1}{2}t^2}$, this is equivalent to showing that, $\lim_{n\to\infty} M_{Z_n}(t) \to e^{\frac{1}{2}t^2}$.

Central Limit Theorem

Proof continued: We need to show that, $\lim_{n\to\infty} M_{Z_n}(t) = e^{\frac{1}{2}t^2}$ or,
$\lim_{n\to\infty}(M(\frac{t}{\sqrt{n}}))^n = e^{\frac{1}{2}t^2}$
But we can take the log on both sides, and this is same as showing,
$\lim_{n\to\infty} n\log(M(\frac{t}{\sqrt{n}})) = \frac{1}{2}t^2$
We will substitute $u = \frac{1}{\sqrt{n}}$, to instead get the easier to deal with limit,
$\lim_{u\to 0} \frac{\log(M(tu))}{u^2}$ and since it is of the form $\frac{0}{0}$, we can use L'Hopital's rule to get,
$= \lim_{u\to 0} \frac{tM'(tu)}{M(tu)2u} = \lim_{u\to 0} \frac{tM'(tu)}{2u}$
We use L'Hopital's rule again to get,
$= \lim_{u\to 0} \frac{t^2M''(tu)}{2} = \frac{1}{2}t^2$

De-Moivre Laplace Theorem

De-Moivre Laplace Theorem

Given a sequence $X_1, X_2, ...,$ of i.i.d random variables with distribution $1(p)$. We define a new sequence of **standardized random variables**, $Z_n = \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}}$. Then the limiting random variable $Z = \lim_{n\to\infty} Z_n$ has the standard normal distribution.

$$Z_n \to N(0, 1)$$

Examples

- Example 1: Flip a fair coin 1000 times. What is $P($ number of heads $> 600)$?
- Example 2: An average of 10 cars arrive every minute at a toll booth. What is $P($ number of cars in the next hour $> 700)$?
- Example 3: The average time a fracture takes to heal is 2 months. What is the probability that the average of healing time of 100 different fractures is less than 1.5 months?

The End